

Apply fine-mapping techniques to summary statistics from a GWAS of Alzheimer's disease in European Population

Team 4

Jianing Wang
Hanqing Zhao

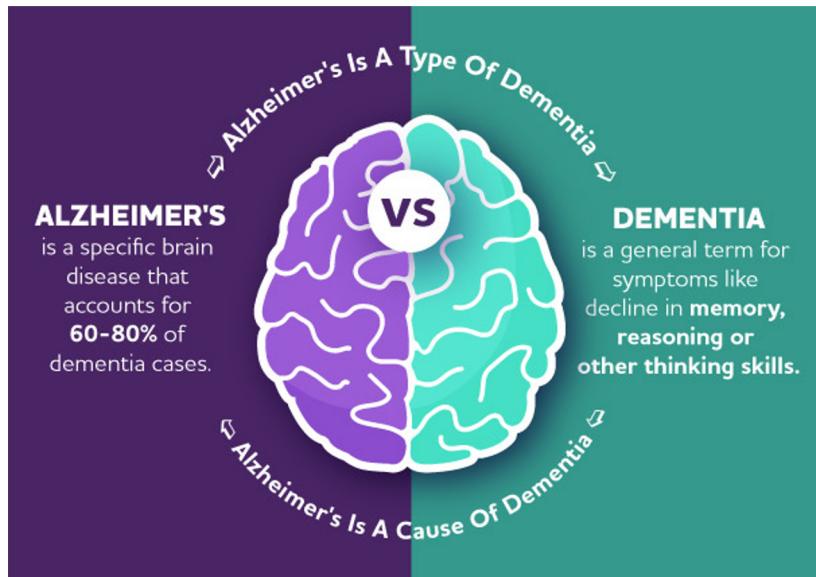
Introduction

Fine-mapping: to refine the genomic relocalization of causal variants by the use of statistical, bioinformatic or functional methods (Schaid *et al.*, 2018).

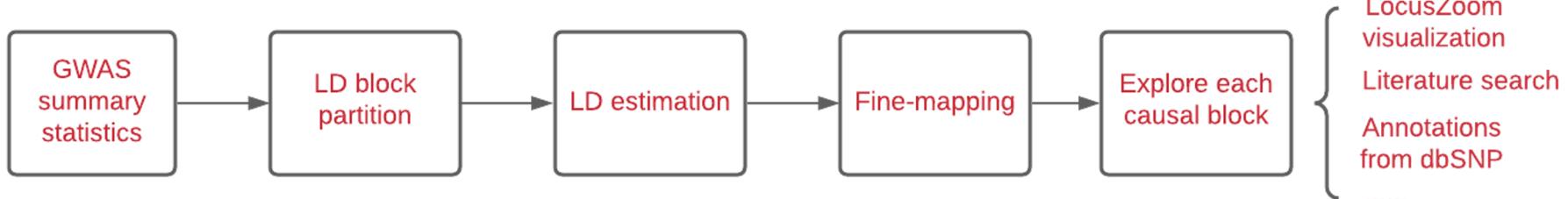
Causal variant: genetic variants that mechanistically contribute to diseases or quantitative traits but are not fully penetrant in the sense that the variant may not be a sufficient cause in isolation (Schaid *et al.*, 2018).

Alzheimer's disease (AD): is a fatal form of dementia and there is currently no cure for this disease. Researches have shown that those who have a close relative with Alzheimer's are more likely to develop the disease, therefore genetics may play a role in developing this disease.

Our project is to see what SNPs might be the causal variants of AD patients in European population by applying fine-mapping tools on the GWAS summary statistic data.



Methods



PAINTOR (Probabilistic Annotation INTEGRATOR) (Kichaev *et al.*, 2014):

- Use the observed marginal test statistics as the true underlying effect size if the observed test statistic is larger than a threshold.
- MCMC algorithms with 1000 Genomes data improves resolution of statistical fine-mapping.

CAVIARBF (CAVIAR Bayes Factor) (Chen *et al.*, 2015):

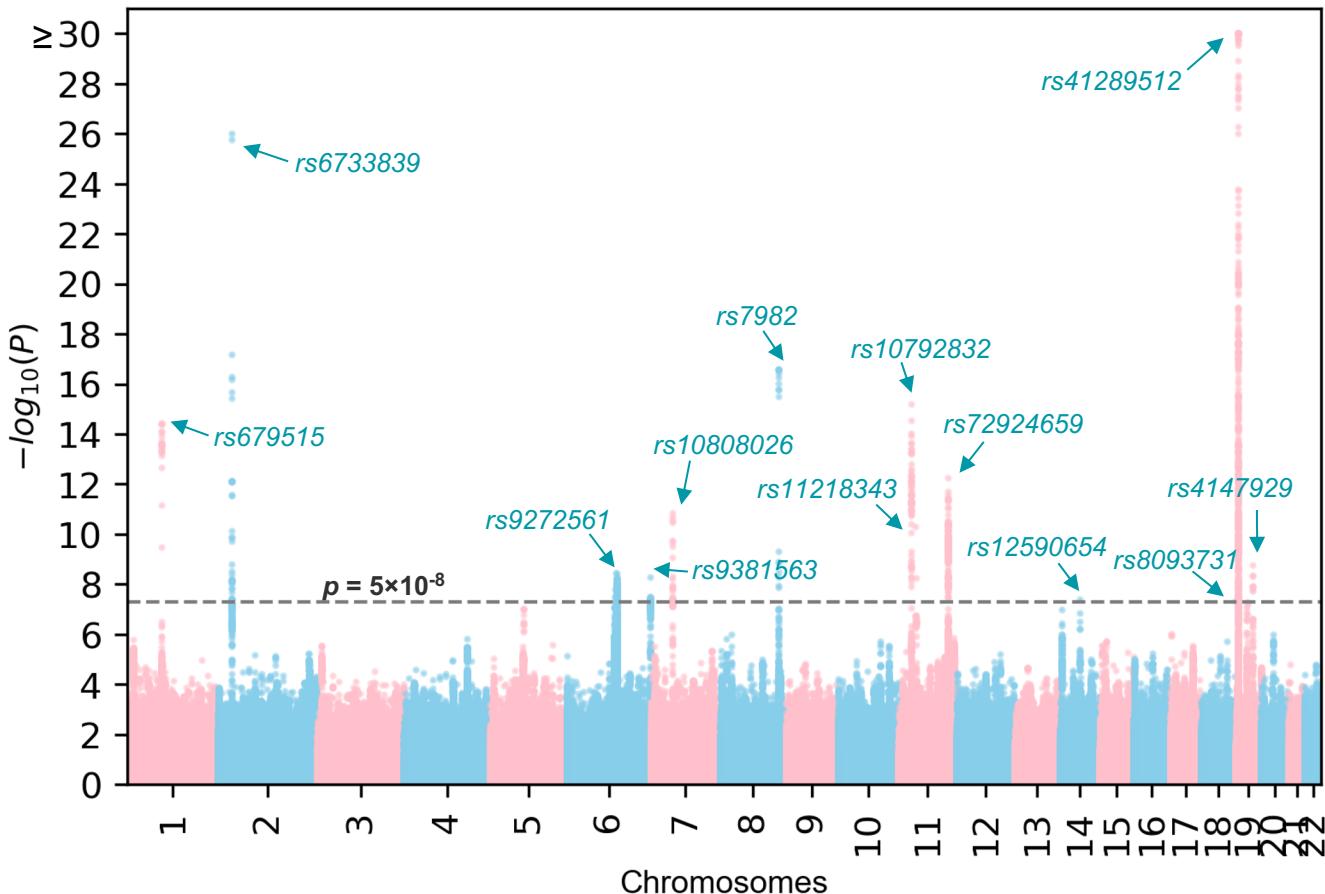
- Use marginal test statistics in the Bayesian framework and the correlation coefficients among SNPs from 1000 Genomes reference panel.
- Model the uncertainty of the effect size (or noncentrality parameters) and perform exhaustive model search

FINEMAP (Benner *et al.*, 2016):

- Use a statistical model similar to CAVIARBF.
- Shotgun Stochastic Search (SSS) algorithms that explore the vast space of causal configurations by concentrating effects on the configurations with non-negligible probability.

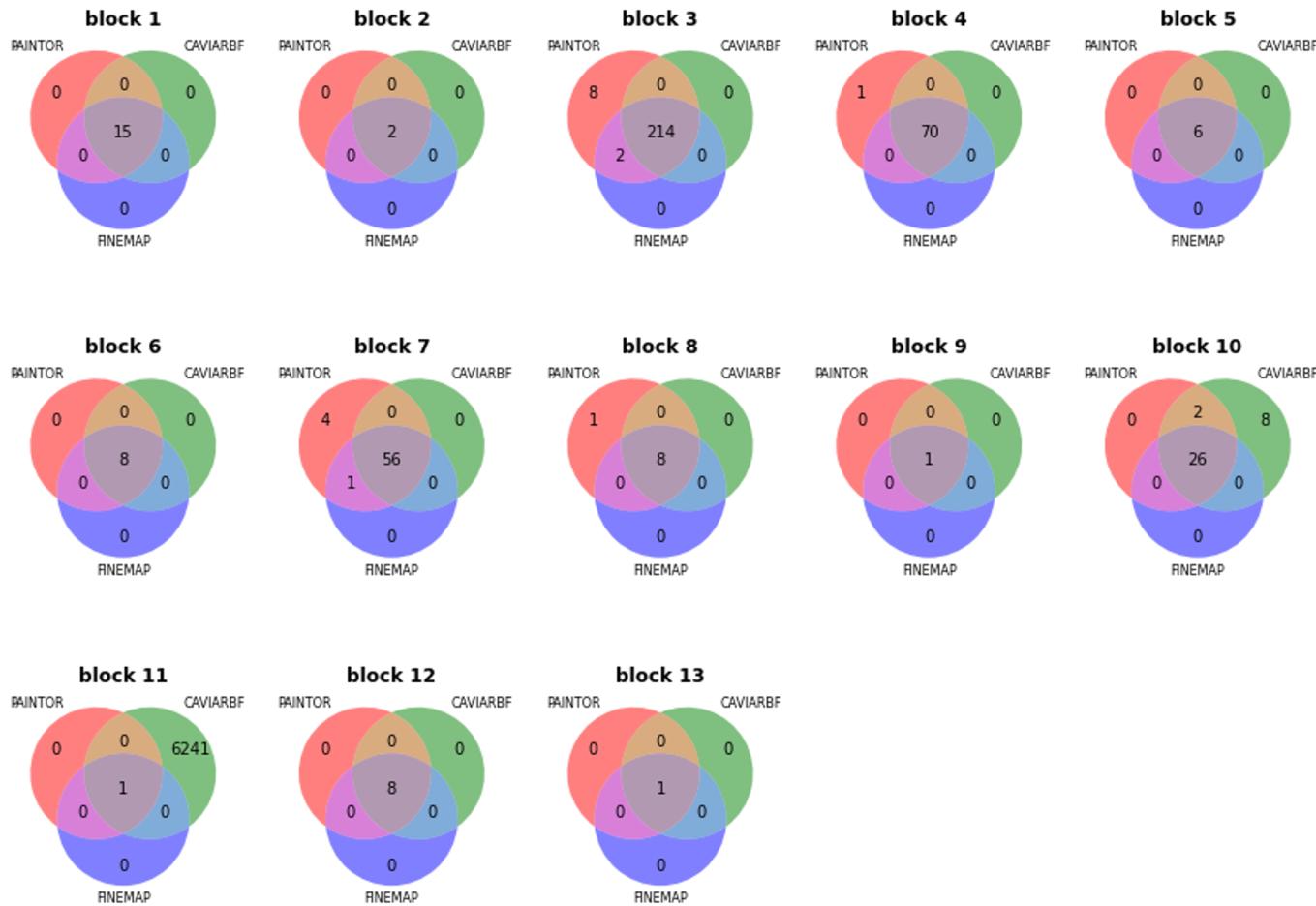
Manhattan plot

- 17,008 cases; 37,154 controls.
- 7,055,881 SNPs in total.
- 13 “skyscrapers”, or **causal blocks**
- the rsIDs of the potential causal variants in each block are labeled on the plot.
- In most cases, the potential causal variant has the lowest GWAS p-values in the block.

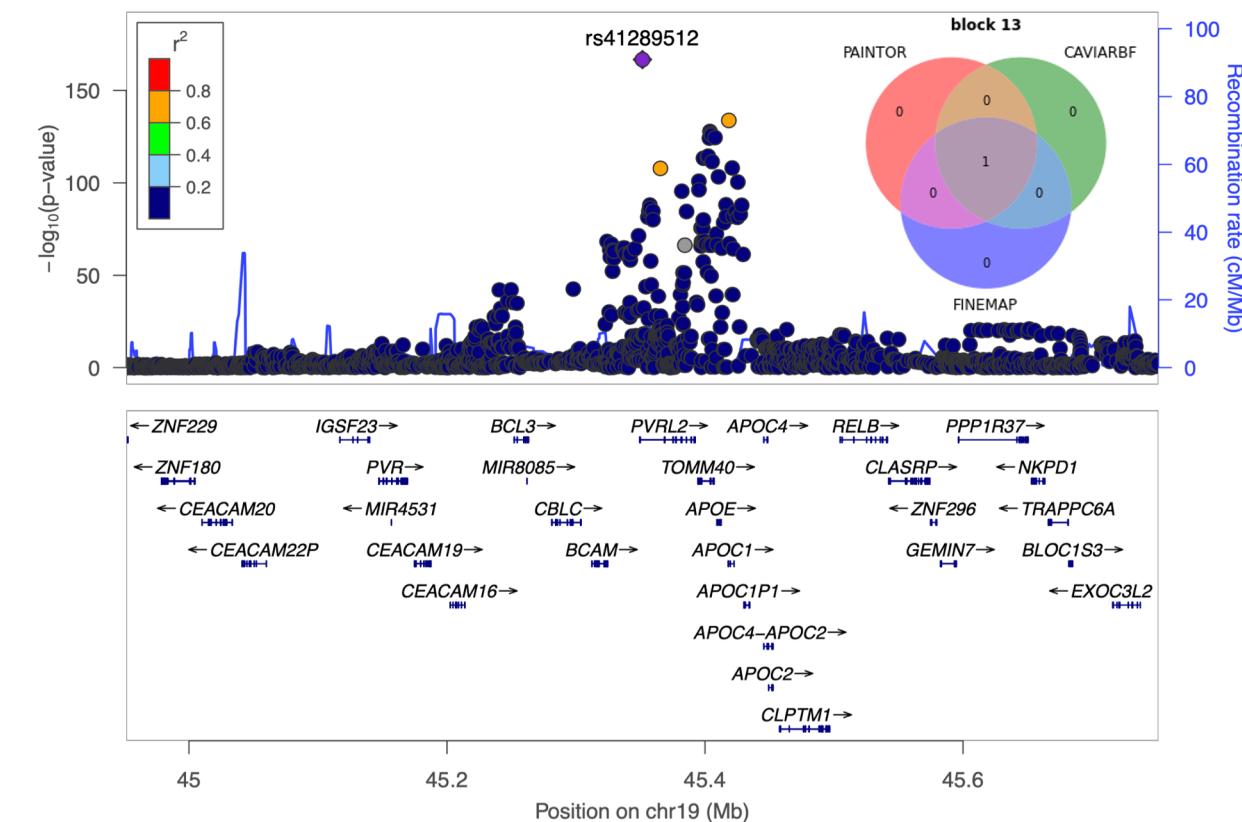


Venn Diagram

- Venn diagrams of the **credible sets** for each causal block.
- *Credible set*: a set of variants with a sum of posterior probabilities of more than α .
- We set the cutoff of credible set to **0.95**.
- In general, results of all three programs are very similar.



CHR	POS	rsID	MAF	REF	ALT	BETA	SE	GWAS p-value			1	2	4	block index		
								P	Zscore	PAINTOR	CAVIARBF	FINEMAP	label	cred_set	gene	annotation
19	45351516	rs41289512	0.0308	G	C	1.6384	0.0594	2.240000e-167	27.582492	1.0	1.0	1.0	7	13	NECTIN2	intron_variant



NECTIN2 / PVRL2:

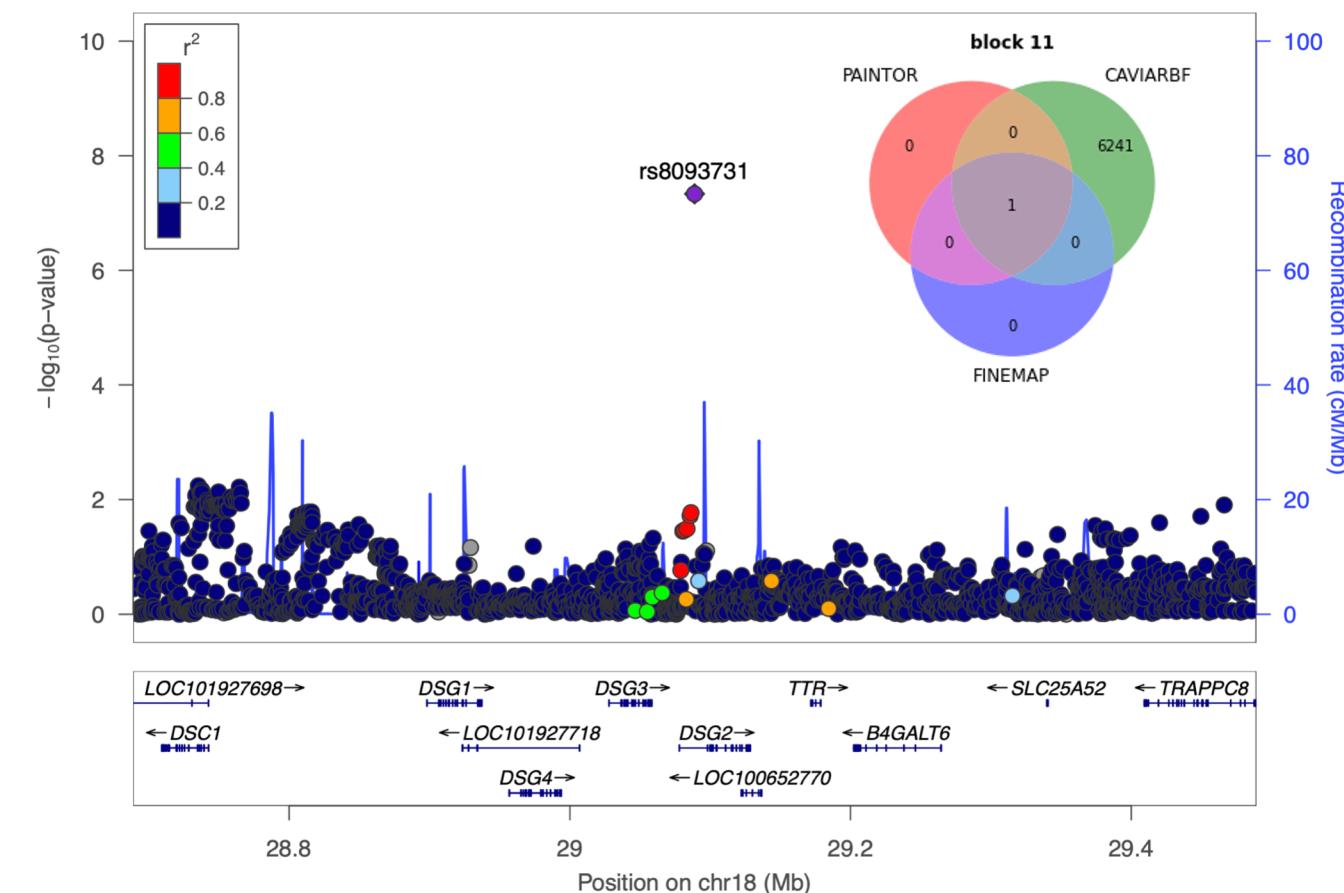
Modulator of T-cell signaling. Can be either a costimulator of T-cell function, or a co-inhibitor, depending on the receptor it binds to.

NECTIN2 is a strong associated genetic variant and also detected in many earlier studies. The potential involvement could be through its role in cell adhesion and brain's susceptibility to viral infections during aging, leading to neuronal loss, and may act together with near associated SNPs such as *APOE* and *TOMM* that are also commonly detected. (Porcellini *et al.*, 2010; Yashin *et al.*, 2017).

membership
to which credible set

dbSNP annotation

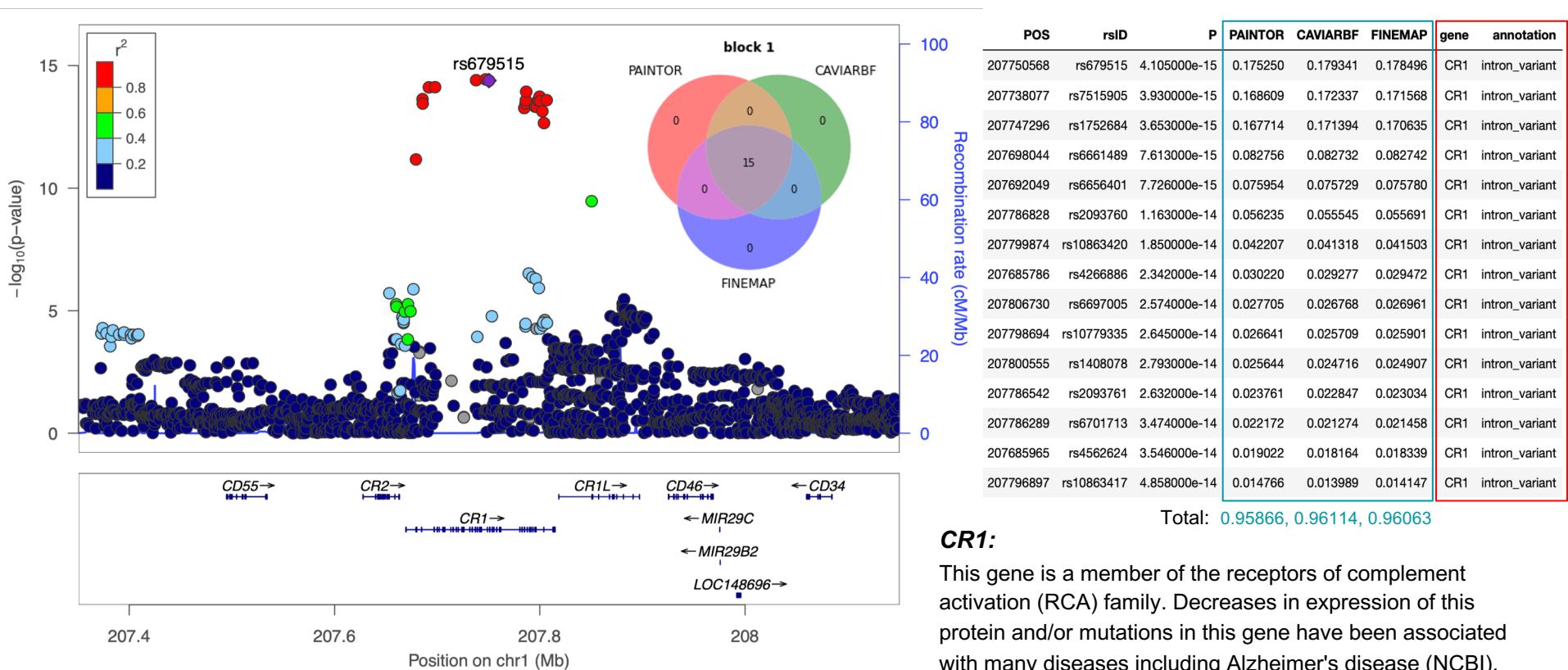
CHR	POS	rsID	MAF	REF	ALT	BETA	SE	P	Zscore	PAINTOR	CAVIARBF	FINEMAP	label	cred_set	gene	annotation
18	29088958	rs8093731	0.0119	T	C	-0.6136	0.1123	4.630000e-08	-5.463936	0.989497	0.934788	0.99224	7	11	DSG2	intron_variant



DSG2:

Component of intercellular desmosome junctions. Involved in the interaction of plaque proteins and intermediate filaments mediating cell-cell adhesion.

DSG2 was previously found to be a risk factor for late-onset AD through GWAS studies, and their functional analysis also suggests DSG2 may be functionally important in the development of APOE ϵ 3/4 allele, which is known to be the strongest genetic risk factor of sporadic AD (Kim *et al.*, 2017).



CR1 has been proved to affect the susceptibility of AD in many previous studies. Those studies found multiple SNPs in *CR1* were significantly linked to amyloid β (A β) metabolism of AD patients, which might be involved in developing AD via regulating A β accumulation. The amyloid β plaques have been regarded as the neuropathological hallmarks of AD (Hardy et al., 2002; Zhu et al., 2020).

Challenges

All three programs require LD data from a reference panel, i.e. 1000 Genome Projects.

However, in PAINTOR's framework, LD computation is VERY time-consuming because the program will go through every VCF file.

Workaround 1: we found a script that can split them into relatively independent LD blocks determined by LDetect and convert them into genotype matrix (Wang, 2020).

Workaround 2: Actually after the computation completed, we found a precomputed reference panel online that uses 1000 Genomes Phase 1 as example.

References

- Porcellini, Elisa et al. "Alzheimer's disease gene signature says: beware of brain viral infections." *Immunity & ageing : I & A* vol. 7 16. 14 Dec. 2010, doi:10.1186/1742-4933-7-16. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3019140/>
- Yashin AI, Fang F, Kovtun M, et al. Hidden heterogeneity in Alzheimer's disease: Insights from genetic association studies and other analyses. *Experimental Gerontology*. 2018 Jul;107:148-160. DOI: 10.1016/j.exger.2017.10.020. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3086666/>
- Reitz, Christiane et al. "Meta-analysis of the association between variants in SORL1 and Alzheimer disease." *Archives of neurology* vol. 68,1 (2011): 99-106. doi:10.1001/archneurol.2010.346. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3086666/>
- Hongwon Kim, Junsang Yoo, Jaein Shin, Yujung Chang, Junghyun Jung, Dong-Gyu Jo, Janghwan Kim, Wonhee Jang, Christopher J Lengner, Byung-Soo Kim, Jongpil Kim, Modelling APOE ε3/4 allele-associated sporadic Alzheimer's disease in an induced neuron, *Brain*, Volume 140, Issue 8, August 2017, Pages 2193–2209, <https://doi.org/10.1093/brain/awx144>
- Zhu, Xc., Dai, Wz. & Ma, T. Impacts of CR1 genetic variants on cerebrospinal fluid and neuroimaging biomarkers in alzheimer's disease. *BMC Med Genet* 21, 181 (2020). [https://doi.org/10.1186/s12881-020-01114-x.](https://doi.org/10.1186/s12881-020-01114-x)
- Hardy J, Selkoe DJ. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science*. 2002 Jul 19;297(5580):353-6. doi: 10.1126/science.1072994. Erratum in: *Science* 2002 Sep 27;297(5590):2209. PMID: 12130773.
- Schaid, D. J., Chen, W., & Larson, N. B. (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nature Reviews Genetics*, 19(8), 491–504. <https://doi.org/10.1038/s41576-018-0016-z>
- Kichaev G, Yang W-Y, Lindstrom S, Hormozdiari F, Eskin E, et al. (2014) Integrating Functional Data to Prioritize Causal Variants in Statistical Fine-Mapping Studies. *PLoS Genet* 10(10): e1004722. doi:10.1371/journal.pgen.1004722
- Chen W, Larrabee BR, Ovsyannikova IG, Kennedy RB, Haralambieva IH, Poland GA, Schaid DJ. Fine Mapping Causal Variants with an Approximate Bayesian Method Using Marginal Test Statistics. *Genetics*. 2015 Jul;200(3):719-36. doi: 10.1534/genetics.115.176107.
- Benner C, Spencer CC, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics*. 2016 May 15;32(10):1493-501. doi: 10.1093/bioinformatics/btw018.
- Jianhua Wang, Dandan Huang, Yao Zhou, Hongcheng Yao, Huanhuan Liu, Sinan Zhai, Chengwei Wu, Zhanye Zheng, Ke Zhao, Zhao Wang, Xianfu Yi, Shijie Zhang, Xiaorong Liu, Zipeng Liu, Kexin Chen, Ying Yu, Pak Chung Sham, Mulin Jun Li, CAUSALdb: a database for disease/trait causal variants identified using summary statistics of genome-wide association studies, *Nucleic Acids Research*, Volume 48, Issue D1, 08 January 2020, Pages D807–D816, <https://doi.org/10.1093/nar/gkz1026>