

FLANDERS (Finemapping coLocalization AND plEiotRopy Solver) pipeline for disentangling pleiotropy of large datasets



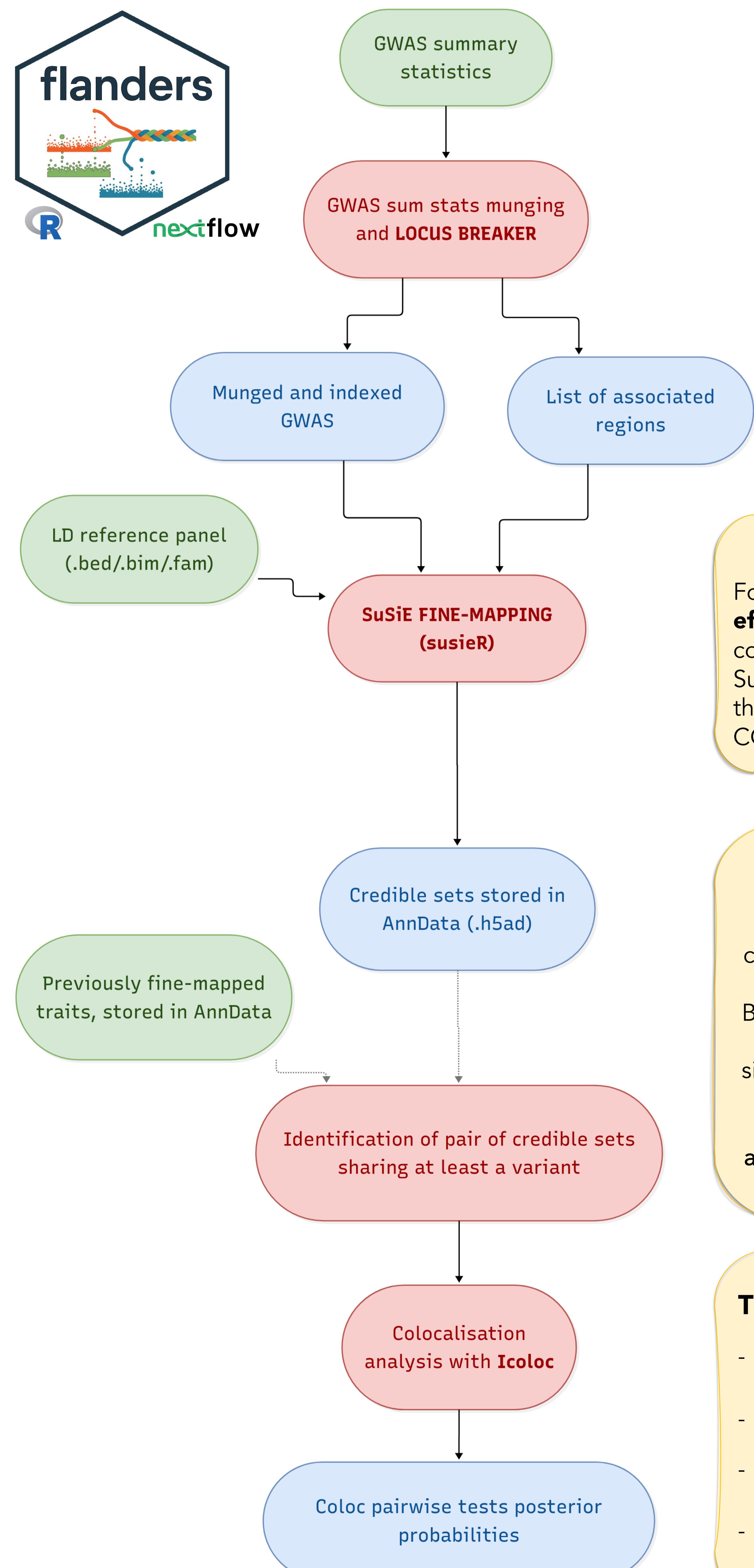
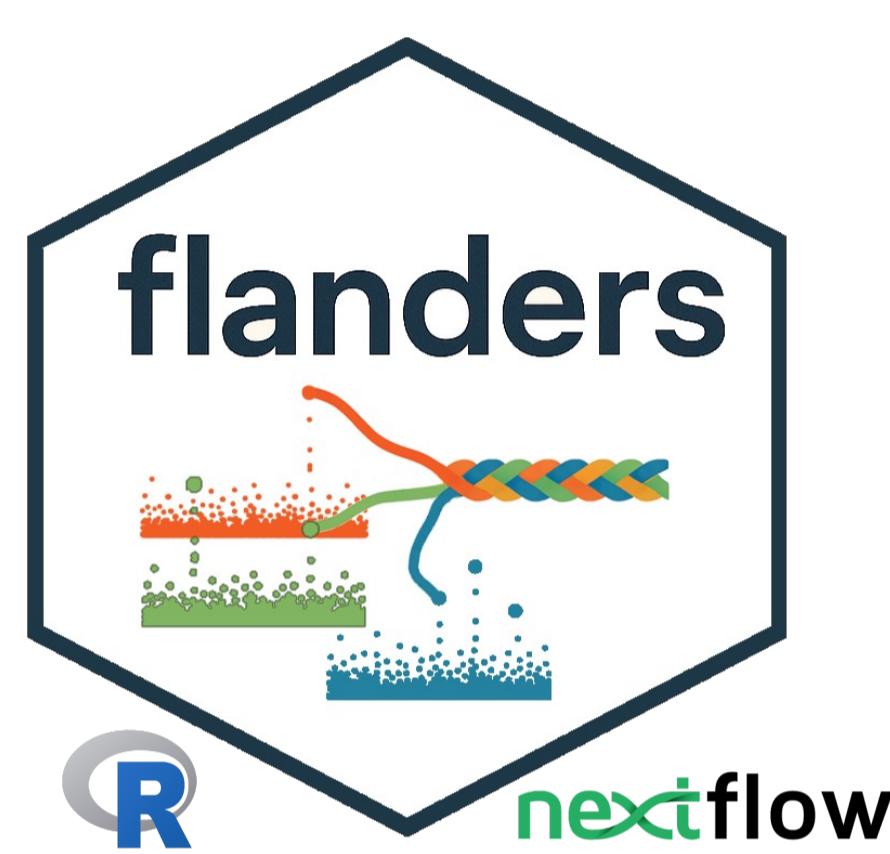
Arianna Landini^{1*}, Sodbo Sharapov^{1*}, Edoardo Giacopuzzi¹, Bruno Ariano¹, Alessandro Raveane¹, Matiss Ozols², Stephen Watt², Cardinal Project Working group, Marc Jan Bonder³, Oliver Stegle^{2,4,5}, Nicole Soranzo^{1,2,6}, Nicola Pirastu¹

¹ Human Technopole, Milan (Italy), ² Wellcome Sanger Institute, Cambridge (UK), ³ Department of Genetics, University of Groningen, Groningen (Netherlands), ⁴ Division of Computational Genomics and Systems Genetics, German Cancer Research Center, Heidelberg (Germany), ⁵ European Molecular Biology Laboratory, Genome Biology Unit, Heidelberg (Germany), ⁶ University of Cambridge, Victor Phillip Dahdaleh Heart & Lung Research Institute, Cambridge (UK). *These authors contributed equally to this work.

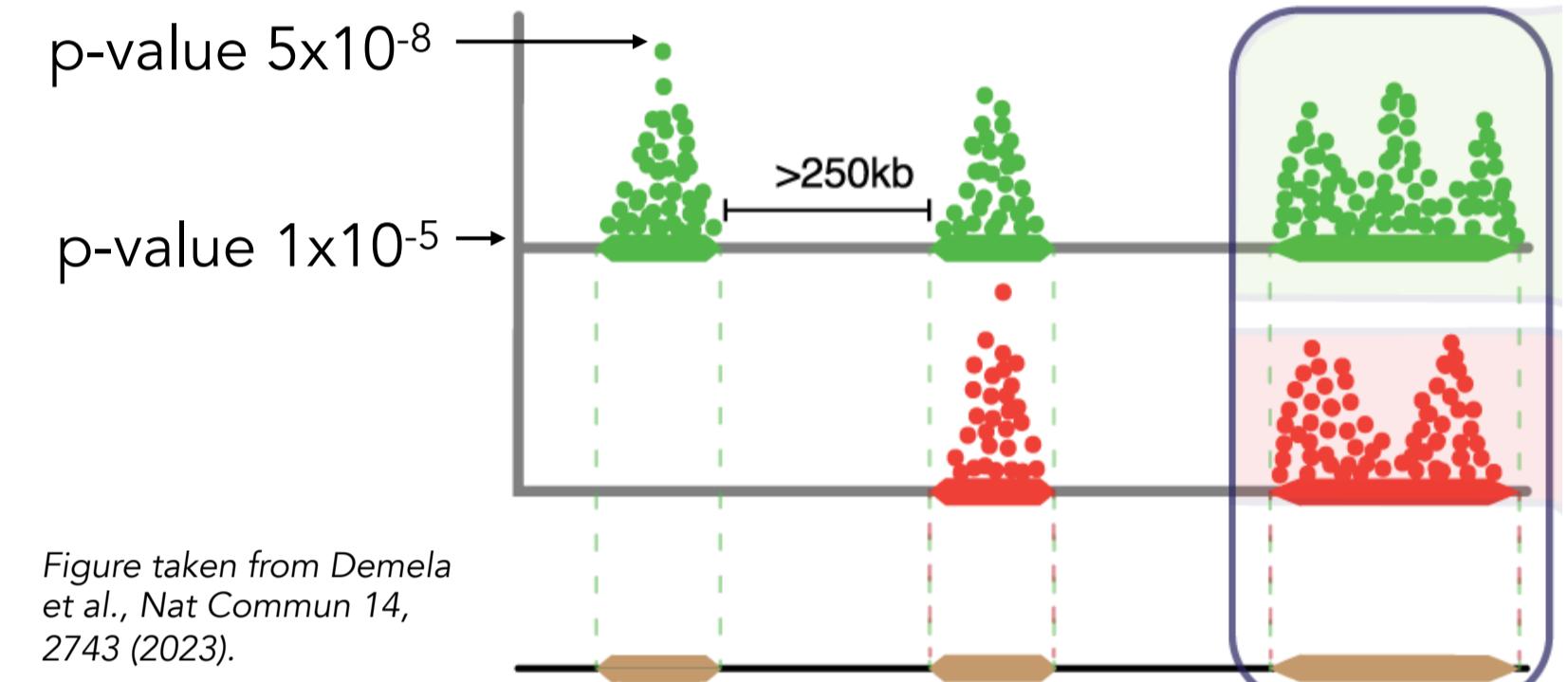
Welcome to Flanders, the fine-mapping and colocalisation pipeline that's as clean and tidy as a Sunday church picnic! This lil' computational crusader takes your big ol' messy GWAS summary, churns out some blessedly tidy credible sets (hallelujah for conditional betas and standard errors!) and wraps it all up in a space-efficient AnnData hug – bless its organized little heart! But wait, there's more! It doesn't just stop at fine mapping – no siree! It hops on over to colocalisation, using Icoloc to impute IABFs smarter than a fox in a theology class.

Howdy-doodly, researchino neighborino!

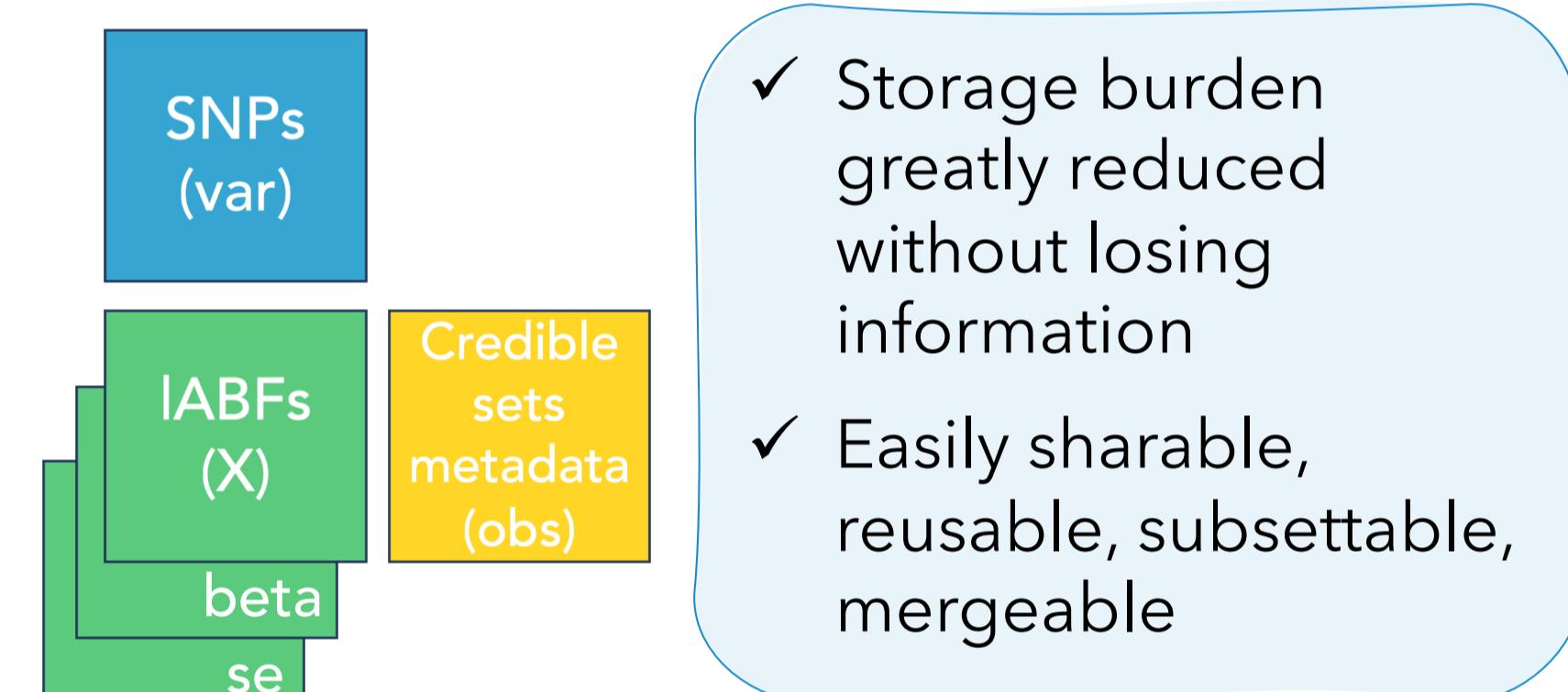
Flanders is here to make your analysis holier, tidier, and oh-so-doodly delightful!



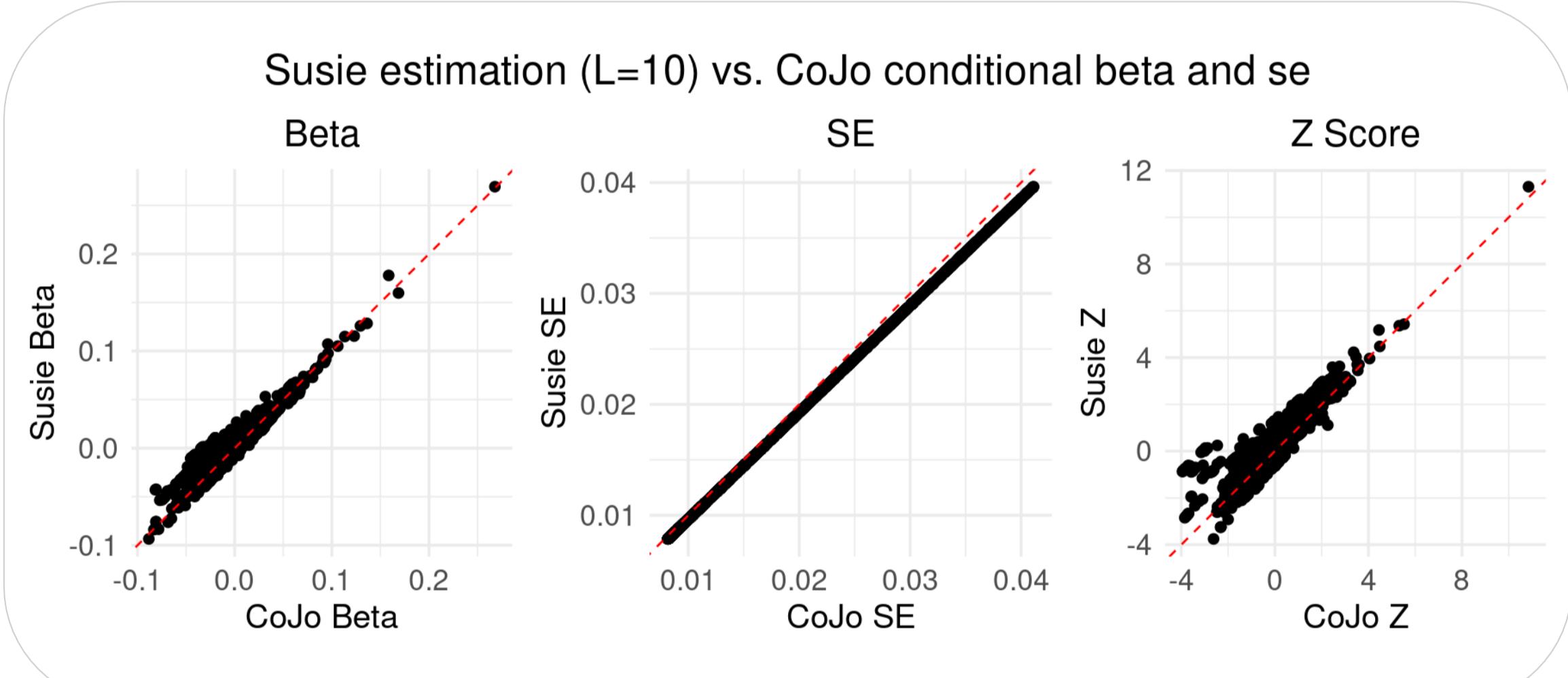
Locus breaker
The algorithm **dynamically defines** the **boundaries** of each significant genomic region by taking into account the shape of the association peak.



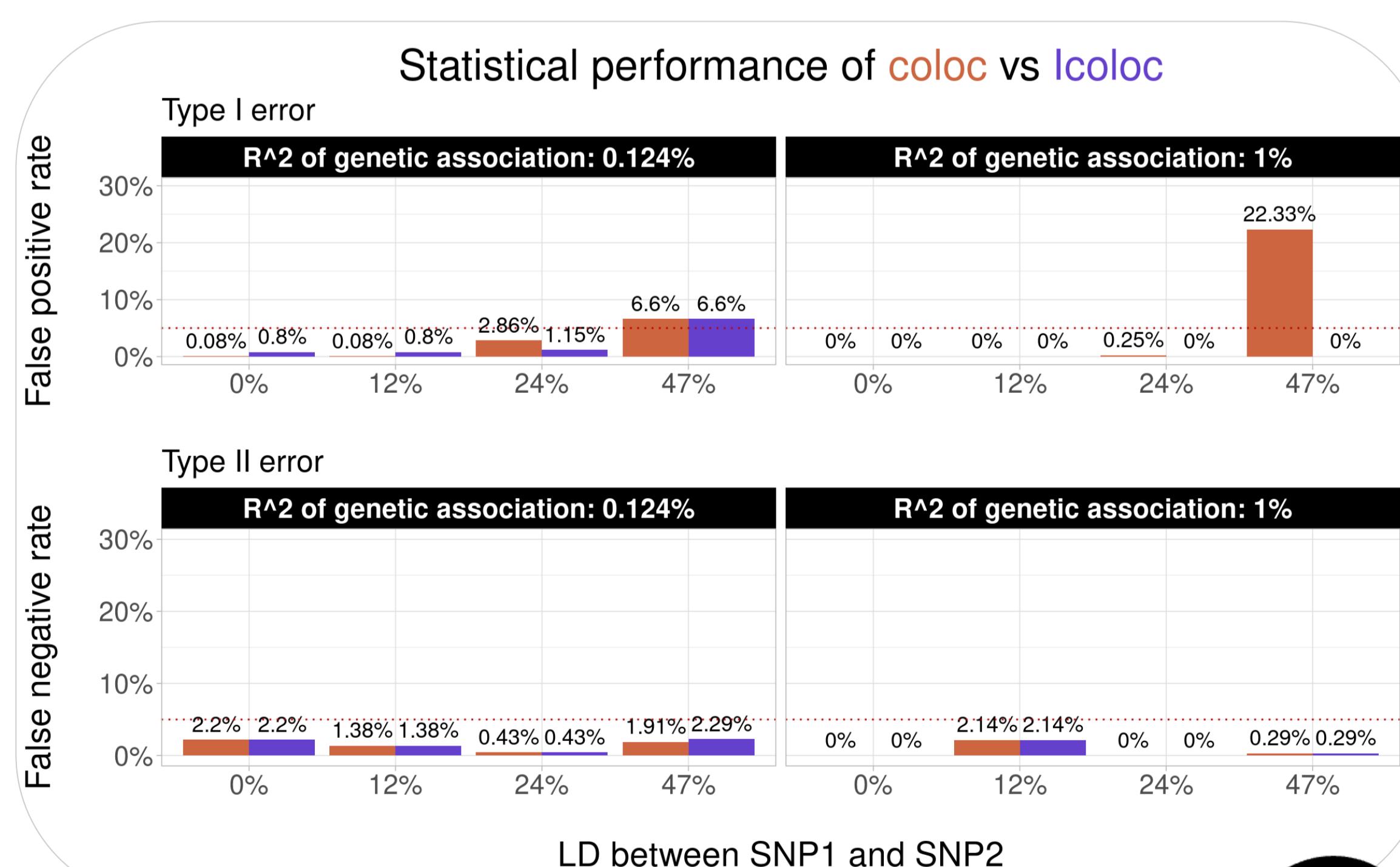
AnnData
Relevant information of each credible set are **efficiently stored** in **AnnData** objects to reduce computational and storage burden.



Conditional effect
For each credible set, **SNP effect estimates** are computed conditionally by SuSiE, largely comparable to those computed by GCTA-COJO.



Icoloc
Extends colocalisation analysis implemented in coloc R package by imputing IABFs (logarithms of the Bayes factor) of SNPs outside of credible sets: this both significantly **reduces storage requirements** but also enhances colocalisation accuracy by **reducing type I error**.



Time and performance in scRNA Cardinal project (UKBB and G&H Pakistani)

- Munging GWAS summary statistics and identifying regions of association for the expression of 15.7K genes in one cell type in the UKBB sample took 33 minutes of real time or 1.16 CPU hours.
- Fine-mapping the 5577 association loci identified took 3 hours of real time or 119 CPU hours.
- Metadata, IABFs, conditional beta and se for ~6.5M SNPs from 98K credible sets (obtained for 39 cell types cis-eQTLs) were stored in ~856MB, reducing disk storage by ~788 times.
- Performing ~518K pair-wise colocalisation tests across credible sets from 39 cell types cis-eQTLs took 10 minutes and 21 seconds of real time or 14.5 CPU hours.

