A. Bootstrap stability vs. resampling — Datasets A–C

K-means Bootstrap Stability vs BOOT_N

A   B   C

Mean Stability (± SD)

Feature set
- genetic
- genetic_fullclin
- genetic_minclin

BOOT_N (resamples)
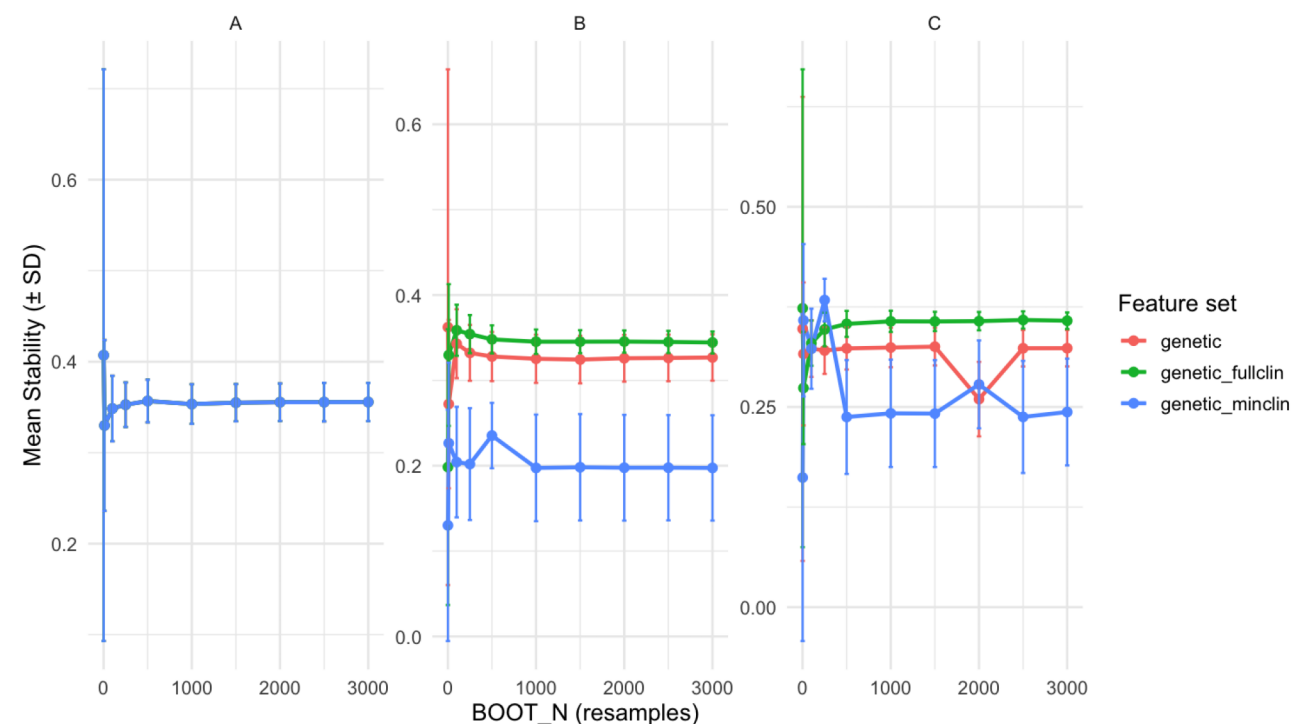
B. Overlap of top-20 SNPs by |PC1 loading|

Top-20 SNPs by |PC1 loading| — feature = genetic

A   B

3 (20%)   0 (0%)   2 (13%)

6 (40%)

1 (7%)   2 (13%)

1 (7%)

C

count
6
5
4
3
2
1
0
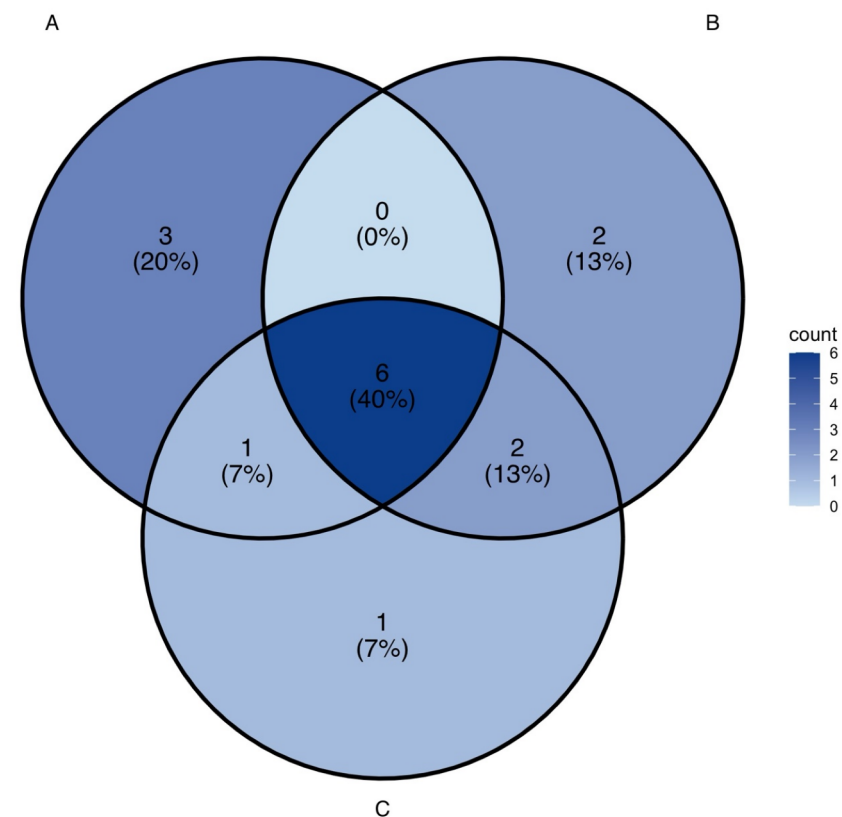
C. Shared SNPs driving genetic separation (Chr6 / HLA locus)

Shared SNPs driving genetic clusters (Chr6 / HLA)

Red points = found among top PC1 loadings in ALL datasets (A, B, C)

|PC1 loading|

06:032615766
06:032616083
06:032644083
06:032644524

06:032655446

Shared A/B/C
- FALSE
- TRUE

Chr6 position (Mb)
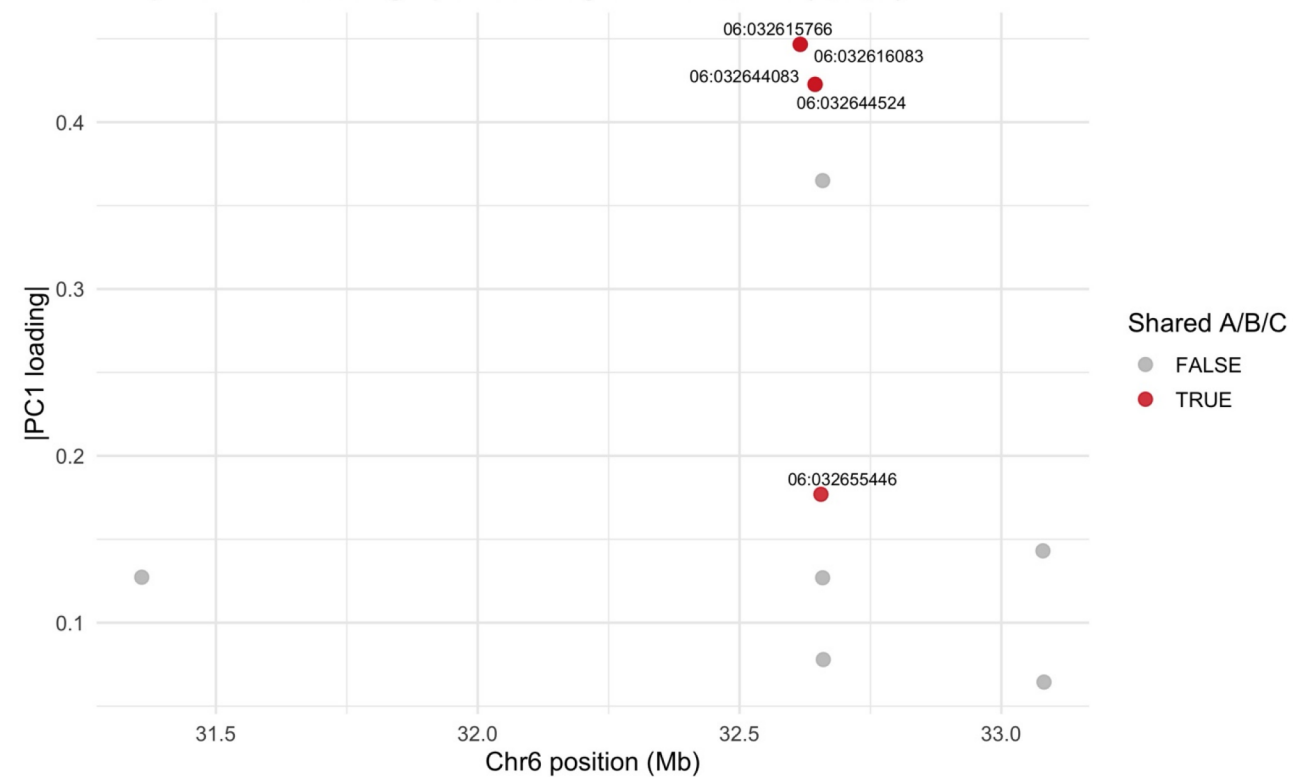
D. Integrated feature model

[Genetic-only model]
• ancestry-driven separation
• low reproducibility

[Clinical-only model]
• phenotype-driven variation
• unstable clusters

[Integrated model]
• combines genotype + phenotype
• stable, reproducible T1D subtypes