Dynamic Rare Variant Association Analysis in Whole Genome Sequencing Studies

Xihong Lin

Harvard T.H. Chan School of Public Health

April 5, 2018

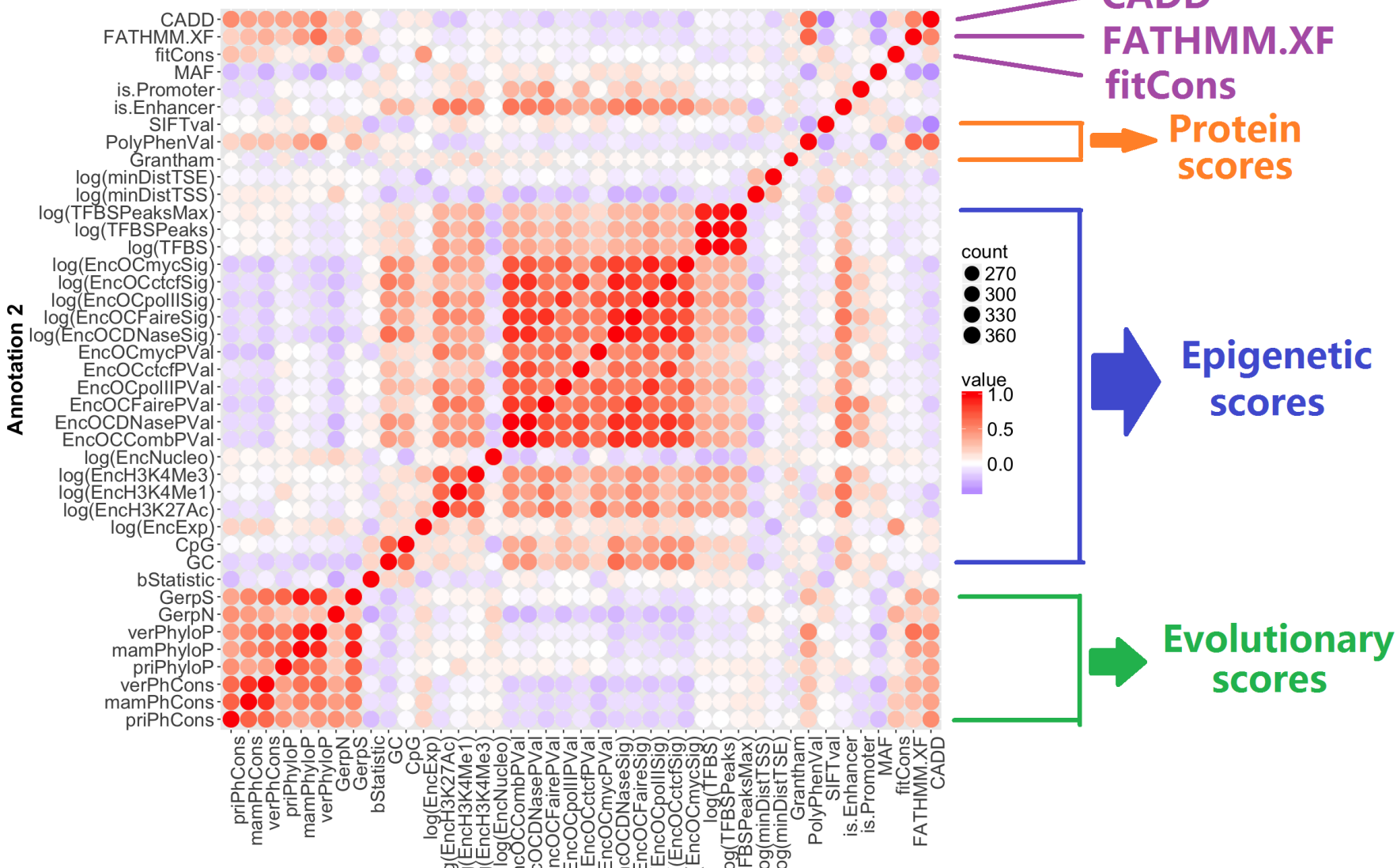
Take Home Message

- Boost power of WGS analysis using dynamic rare variant association methods by
  - SKAT-C/Burden-C: weighting variants using annotations optimally determined by the data for any given SNV set;
  - SCANG: scanning the genome using dynamic window sizes to detect association segment locations and sizes.

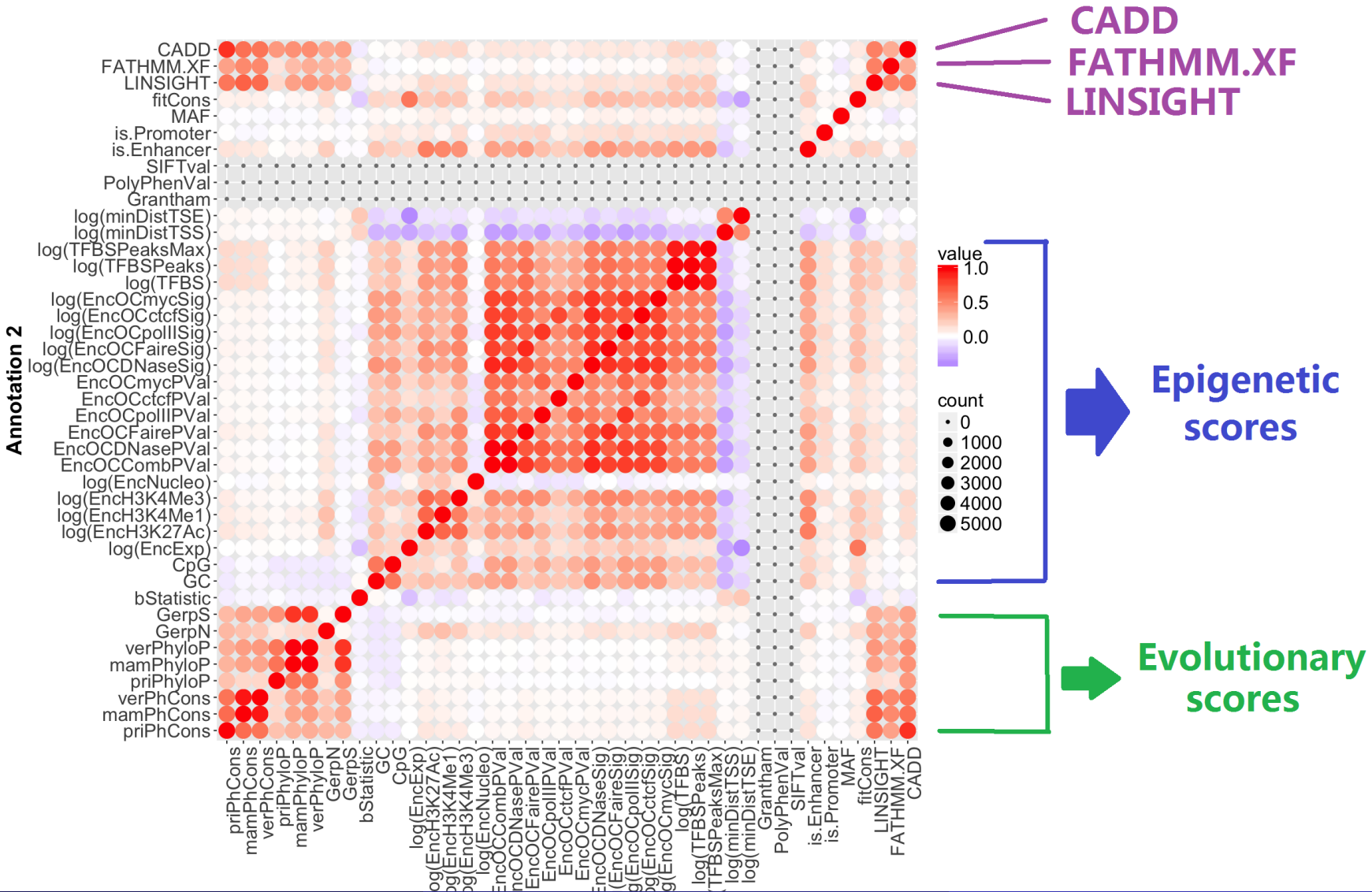
Challenges in WGS Rare Variant (RV) Analysis

- **SNV-Set analysis:** Individual SNV analysis has little power
- **Choices of SNV-Sets:**
  - Gene-centric analysis: different functional categories of genes, e.g, LOF, missense, promoters, enhancers,
  - Moving window : a pre-fixed window size
  - Dynamic scan:SCAN Genome by data driven optimal window sizes (**SCANG**)

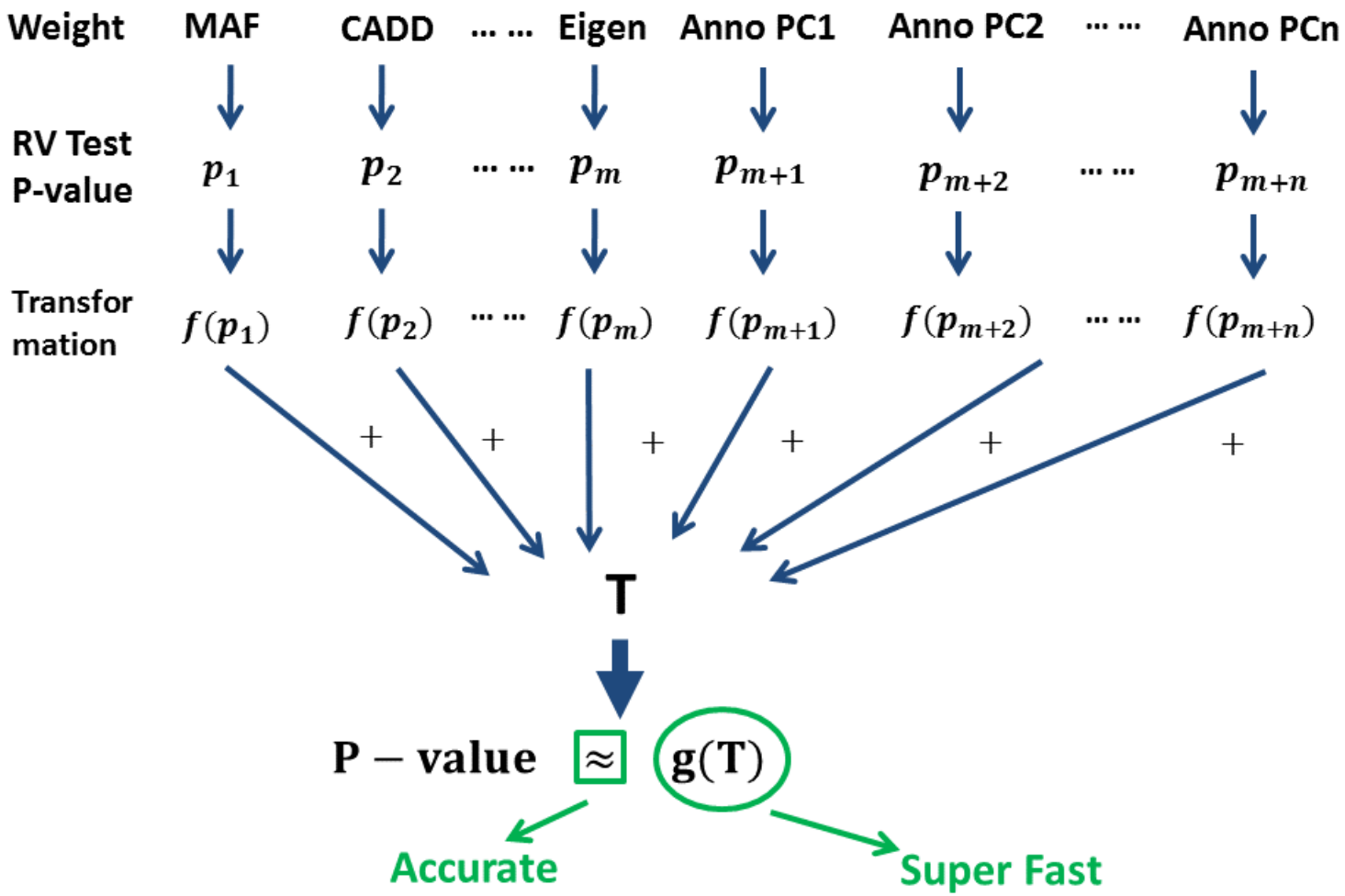
Correlation Heatmap of Functional Scores: Coding Variants



Correlation Heatmap of Functional Scores: Noncoding Variants



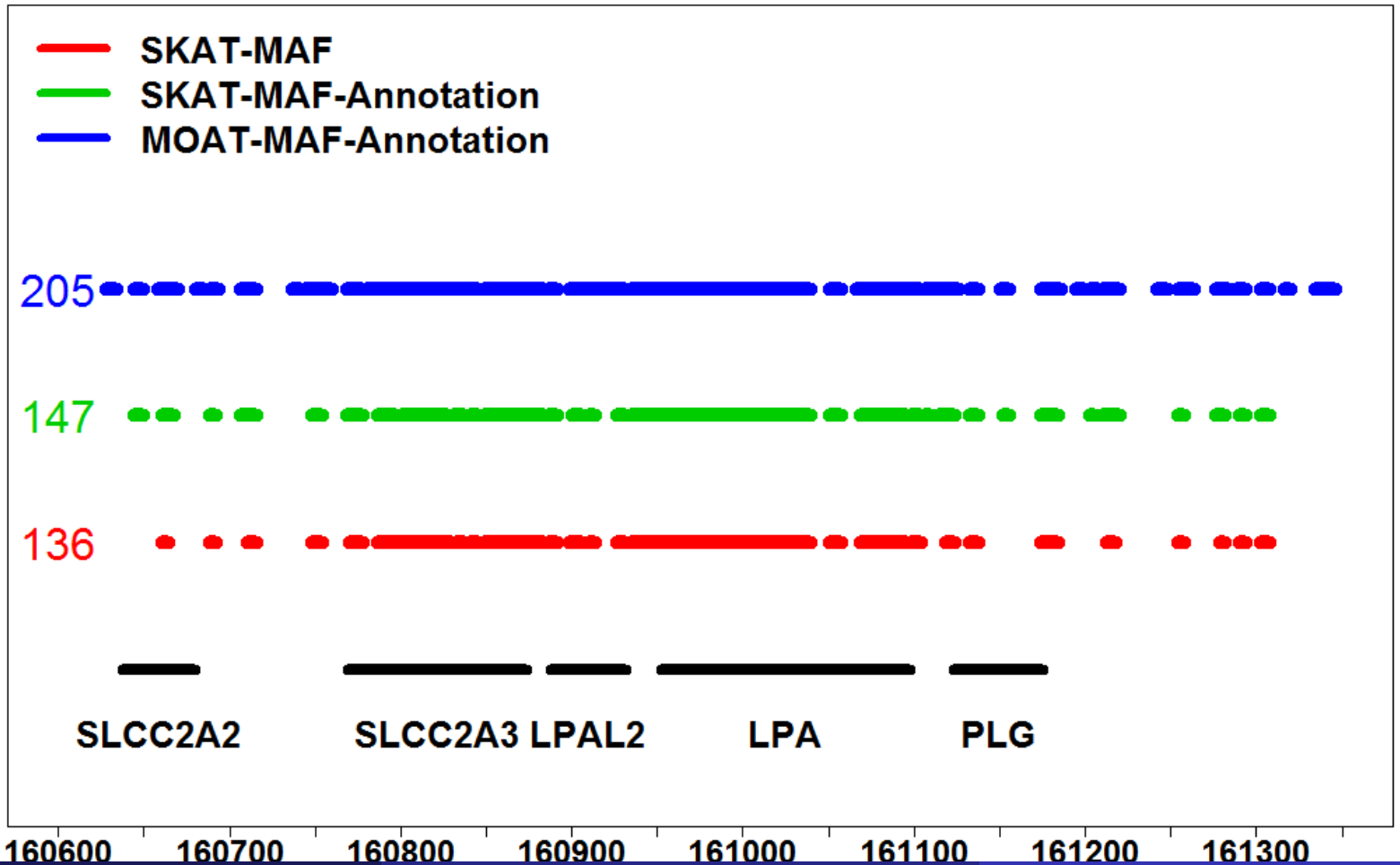
SKAT-C/Burden-C (Optimally weighted by annotations)



Whole Genome Sequencing Analysis: ARIC Data

- Acknowledgment: Eric Boerwinkle
- Phenotype: Lpa (lipoprotein) and LDL.
- Sample sizes: $\sim$ 1800 AA (55M SNVs) and $\sim$ 1400 EA (31M SNVs)
- Covariates: age, sex and the first 3 PCs.
- Rare and low-frequent variants (MAF < 0.05)

Lpa(AA): Significant 4KB Sliding Windows in Chr 6 ($p < 3.75 \times 10^{-8}$) Adjusting for rs10455872 and rs3798220(Surrogates of Kringle Type IV Repeats)

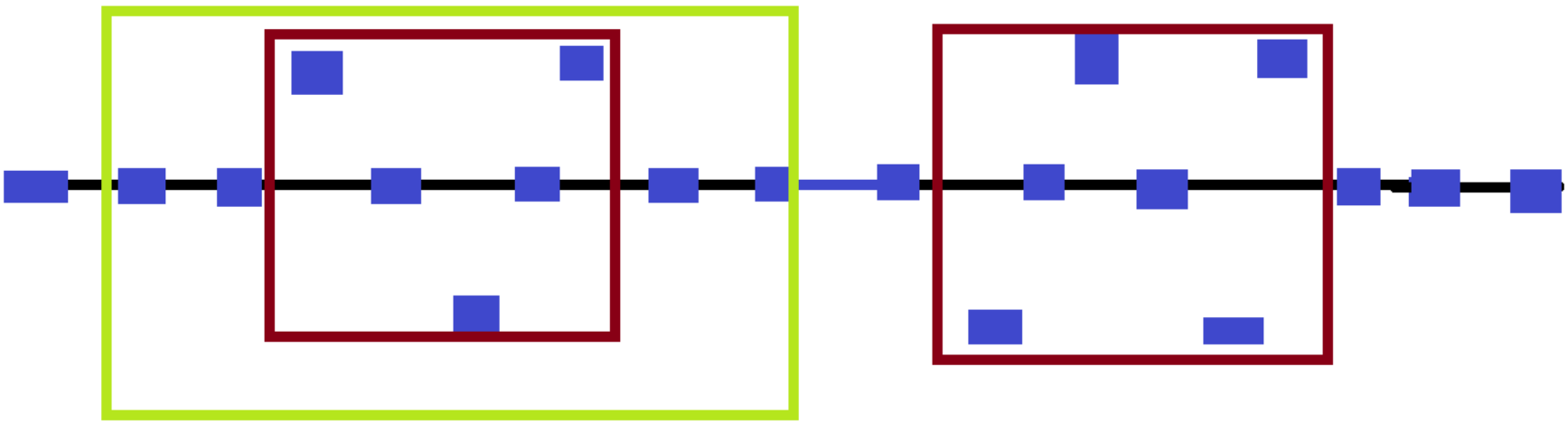


SCANG: SCAN the Genome

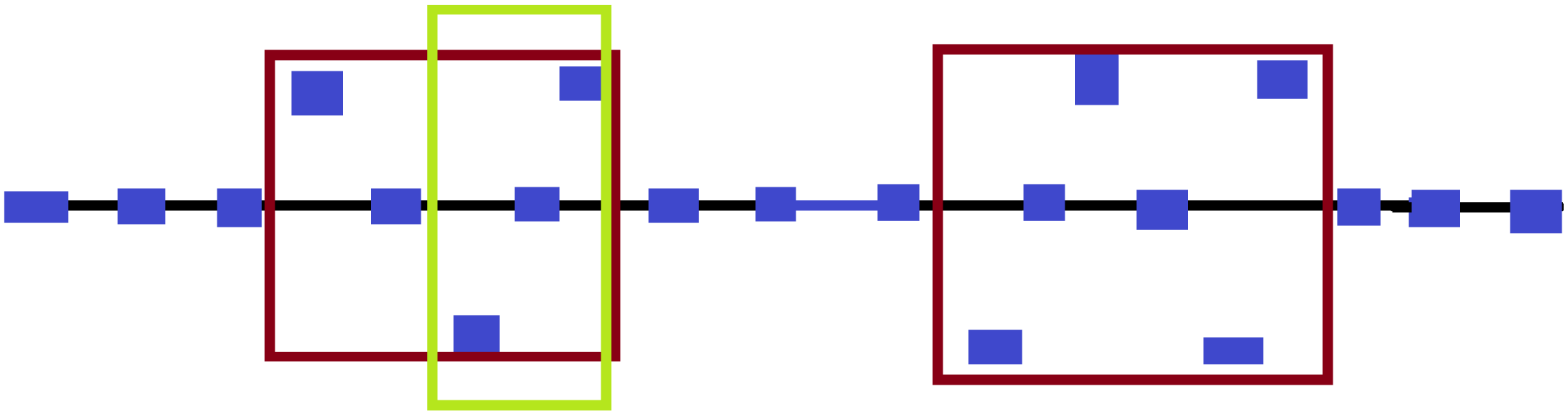
Key Features:

- Use to the data to flexibly detect the RV association signal region sizes and locations without a prior fixed window size
- Control the **genome-wide** type I error rate at $\alpha = 5\%$

Sliding Window is Too Big

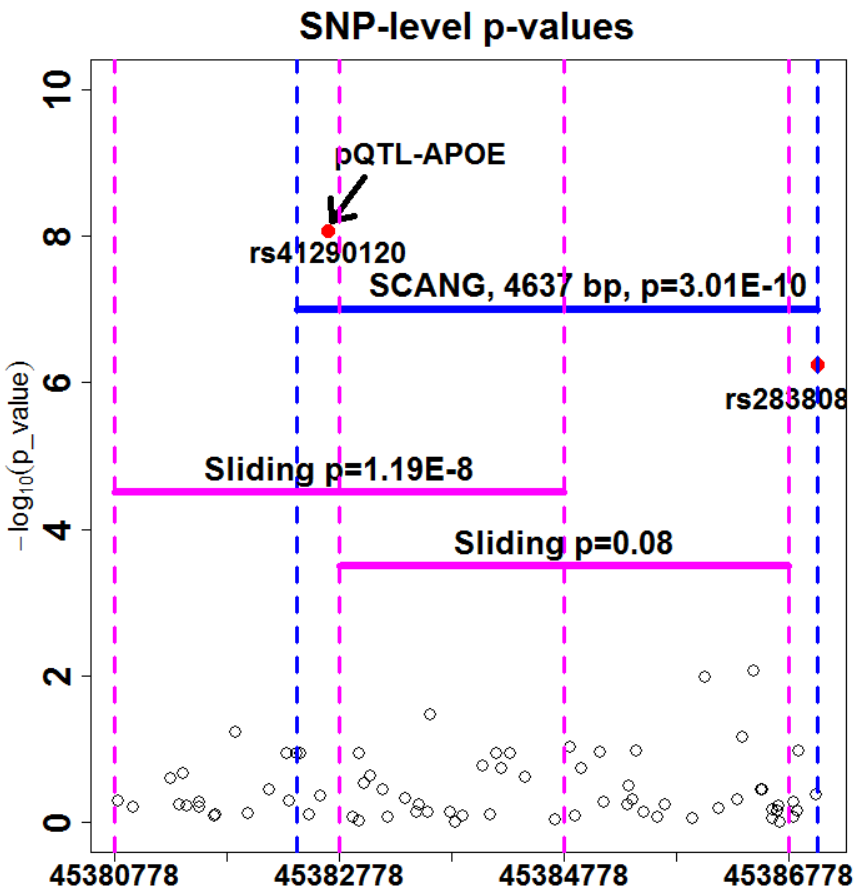


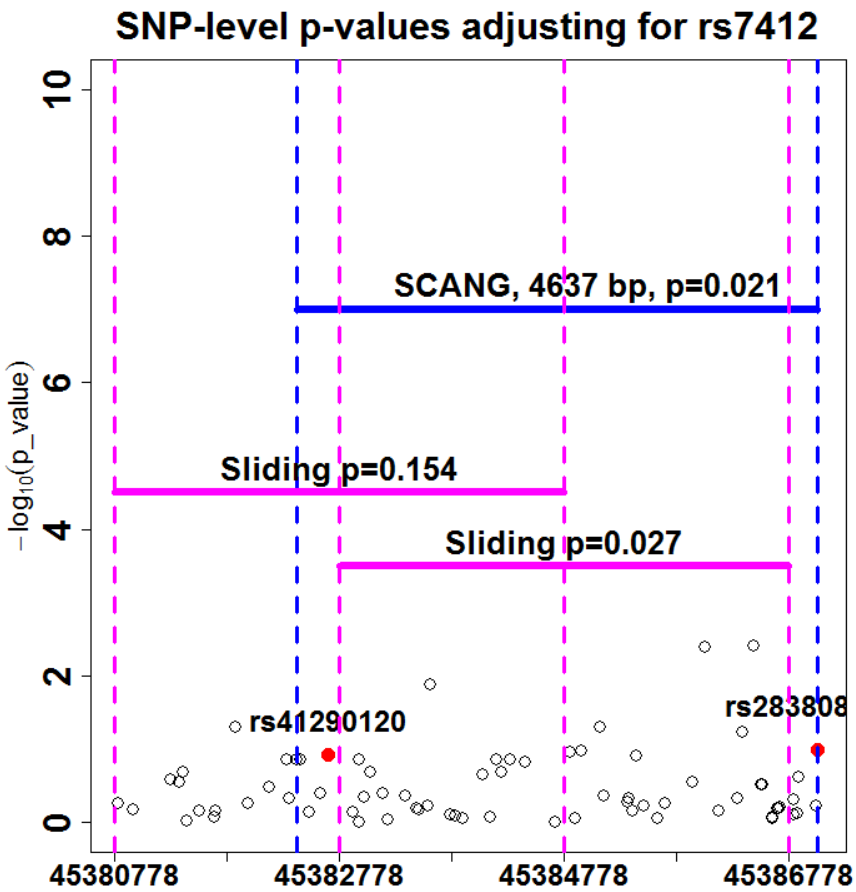
Sliding Window is Too Small



Statistical goal: Detect the red signal box.

Comparison of 4KB Sliding Window and SCANG: PVRL2 for LDL in EA





Concluding Remarks

- New GSP and TOPMed Annoation WGs: To collaborate on building comprehensive and consistent annotations (Annotation WGs)
- RV analysis accounting for relatedness: AMMAT
- Develop the cloud-version of the new association methods for HAIL(GSP), EPACTS (TOPMed), Analysis Commons (TOPMed).