Recent and Future Updates to LDpred2 for Polygenic Scores and Inference

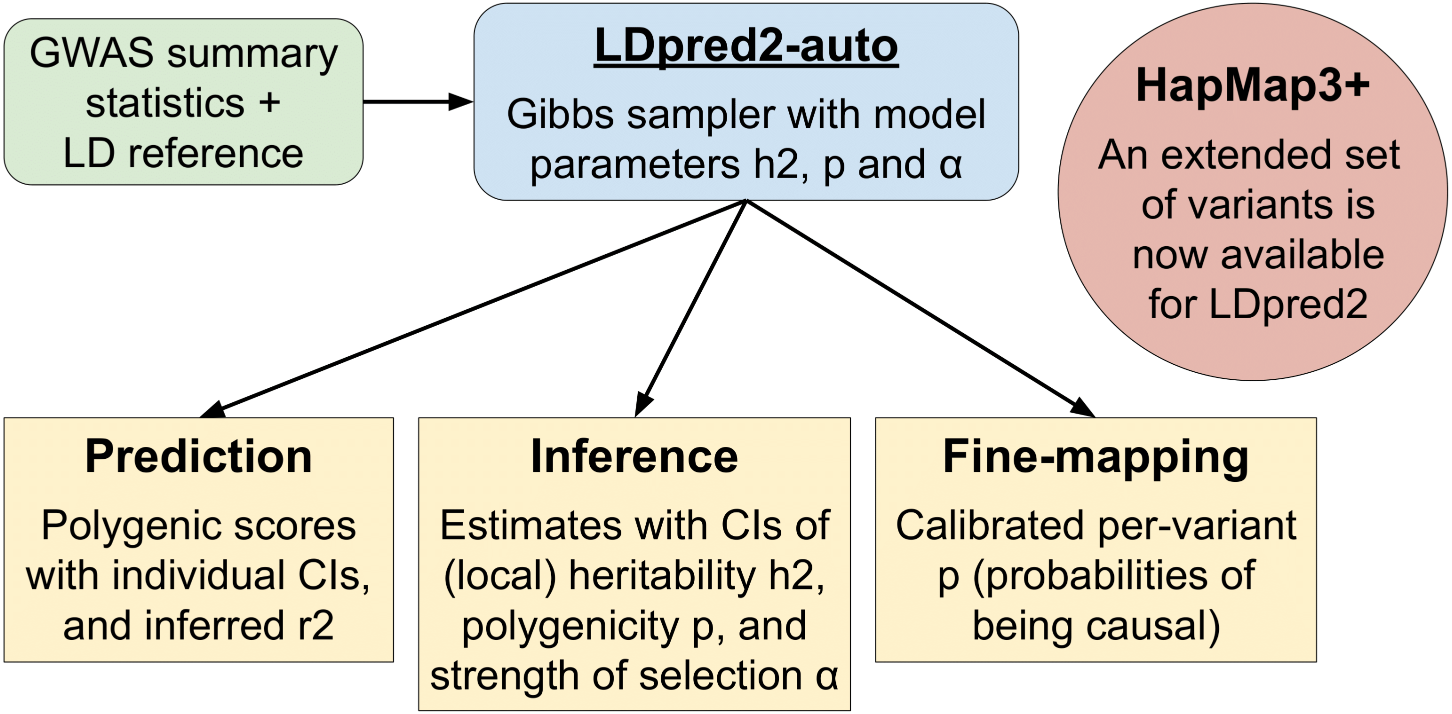
WCPG 2023

Florian Privé (Aarhus Uni, DK)

privefl

Nothing to disclose

Overview of what LDpred2-auto can now provide



GWAS summary statistics + LD reference

LDpred2-auto

Gibbs sampler with model parameters h2, p and α

HapMap3+

An extended set of variants is now available for LDpred2

Prediction

Polygenic scores with individual CIs, and inferred r2

Inference

Estimates with CIs of (local) heritability h2, polygenicity p, and strength of selection α

Fine-mapping

Calibrated per-variant p (probabilities of being causal)

Prior model: spike and slab

LDpred2 assumes the following model for effect sizes,

$$\beta_j = S_j \gamma_j \sim \begin{cases} \mathcal{N}\left(0, \dfrac{h^2}{Mp}\right) & \text{with probability } p, \\ 0 & \text{otherwise,} \end{cases}$$

where

- p is the proportion of causal variants (aka polygenicity),
- M the number of variants,
- h^2 the (SNP) heritability,
- $\boldsymbol{\gamma}$ the effect sizes on the allele scale,
- $\boldsymbol{S}$ the standard deviations of the genotypes,
- $\boldsymbol{\beta}$ the effects of the scaled genotypes.

Prior model: spike and slab

LDpred2 assumes the following model for effect sizes,

$$\beta_j = S_j \gamma_j \sim \begin{cases} \mathcal{N}\left(0, \dfrac{h^2}{Mp}\right) & \text{with probability } p, \\ 0 & \text{otherwise,} \end{cases}$$

where

- p is the proportion of causal variants (aka polygenicity),
- M the number of variants,
- h^2 the (SNP) heritability,
- $\boldsymbol{\gamma}$ the effect sizes on the allele scale,
- $\boldsymbol{S}$ the standard deviations of the genotypes,
- $\boldsymbol{\beta}$ the effects of the scaled genotypes.

LDpred2 uses a Gibbs sampler to sample causal variants and their effects.
LDpred2-auto directly estimates h^2 and p from the Gibbs sampler.

Extended 3-parameter model (for LDpred2-auto)

$$\beta_j = S_j \gamma_j \sim \begin{cases} \mathcal{N}\left(0,\ \sigma_\beta^2 \cdot (S_j^2)^{(\alpha+1)}\right) & \text{with probability } p, \\ 0 & \text{otherwise.} \end{cases}$$

- similar to the model assumed in SBayesS
- previous 2-parameter model assumes $\alpha = -1$ and $\sigma_\beta^2 = \frac{h^2}{Mp}$
- σ_β^2 and α are estimated using maximum likelihood estimation (MLE)

Extended 3-parameter model (for LDpred2-auto)

$$\beta_j = S_j \gamma_j \sim \begin{cases} \mathcal{N}\left(0,\ \sigma_\beta^2 \cdot (S_j^2)^{(\alpha+1)}\right) & \text{with probability } p, \\ 0 & \text{otherwise.} \end{cases}$$

- similar to the model assumed in SBayesS
- previous 2-parameter model assumes $\alpha = -1$ and $\sigma_\beta^2 = \frac{h^2}{Mp}$
- σ_β^2 and α are estimated using maximum likelihood estimation (MLE)

⚠ More flexibility is not always better.
Inference can become unstable when power is low.
The 2-parameter model can be used instead by setting `use_MLE = FALSE`.

Inference with LDpred2-auto

Recent work on:

- properly validating the inference of h^2, p, and α (and their CIs) using extensive simulations
- showing calibrated per-variant probabilities of being causal
- inferring the out-of-sample predictive performance r^2 directly from the Gibbs sampler

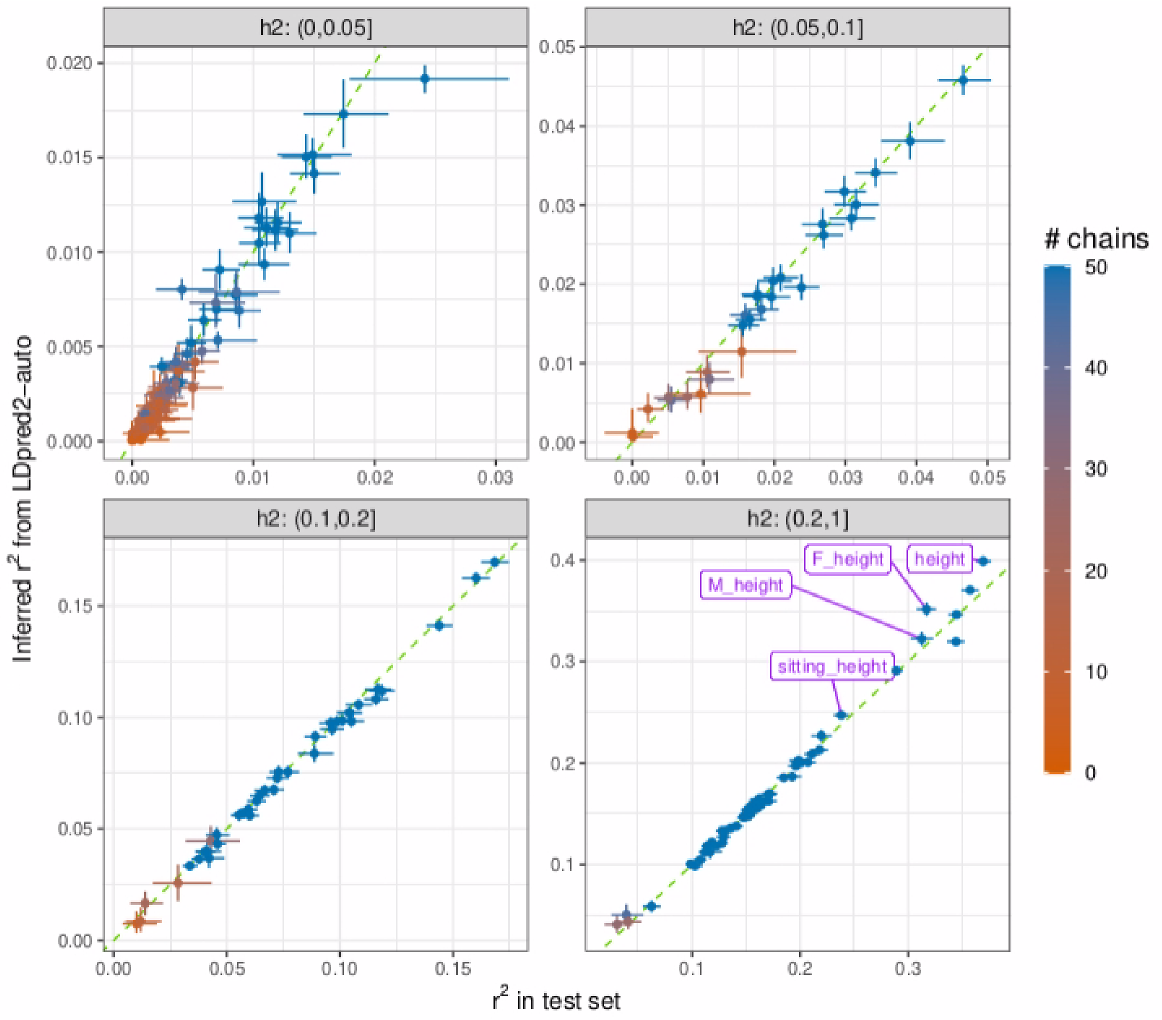
Inference with LDpred2-auto

Recent work on:

- properly validating the inference of h^2, p, and α (and their CIs) using extensive simulations
- showing calibrated per-variant probabilities of being causal
- inferring the out-of-sample predictive performance r^2 directly from the Gibbs sampler

ℹ Soon to be published in AJHG. Stay tuned @privefl.

Inferred r^2 estimates vs the ones from a test set



An extended set of variants for LDpred2

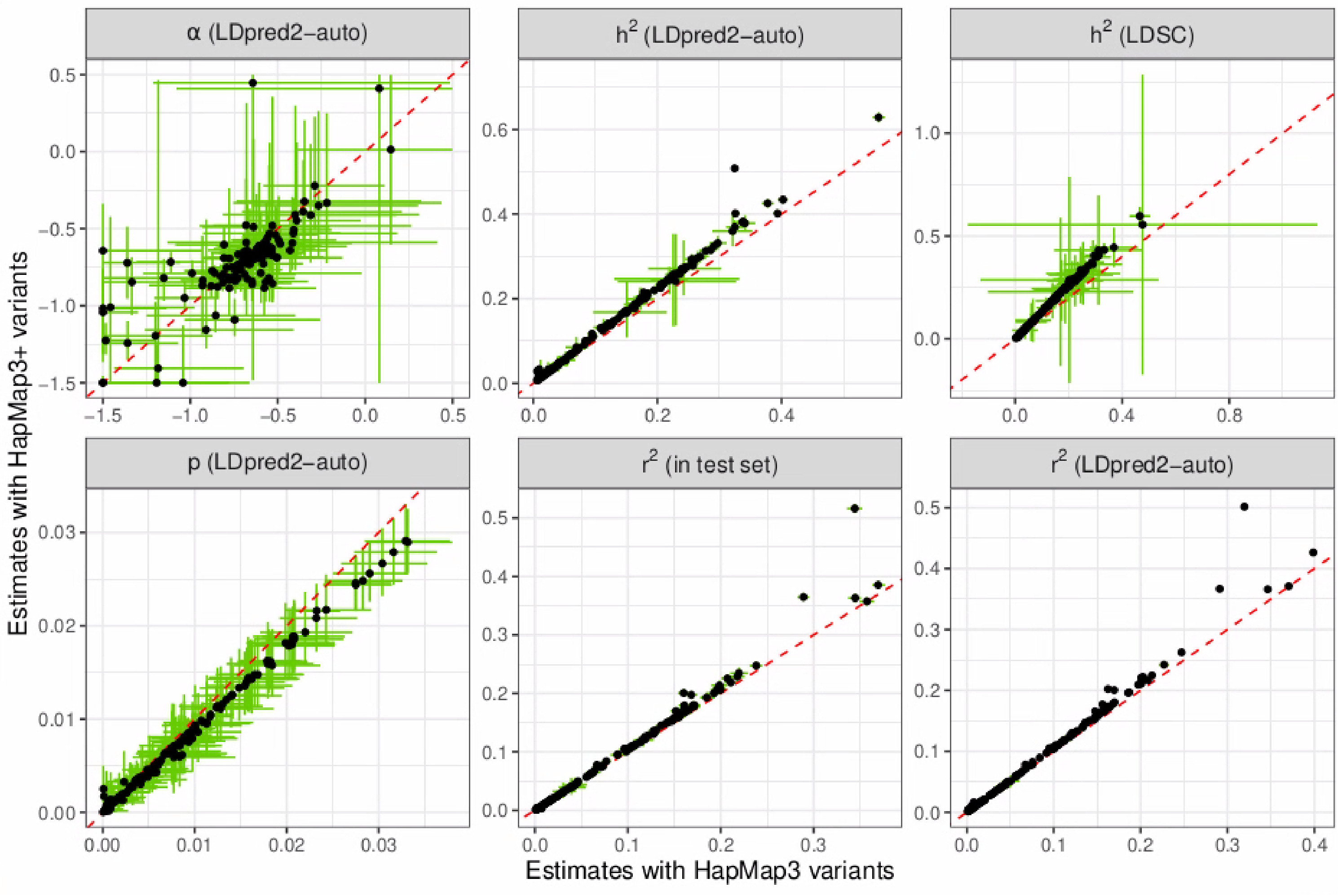
- We have recommended to use a set of 1,054,330 HapMap3 variants
  - good coverage of the genome
  - generally well imputed and available in most studies
- We now provide an extended set with 37% more variants
  - designed to maximize tagging of 11.5M common variants in diverse genetic ancestries
  - called HapMap3+
- Using this new set of variants, in UK Biobank analyses, *on average*,
  - we capture 12% more SNP heritability h^2
  - obtain 6% more predictive performance r^2

An extended set of variants for LDpred2

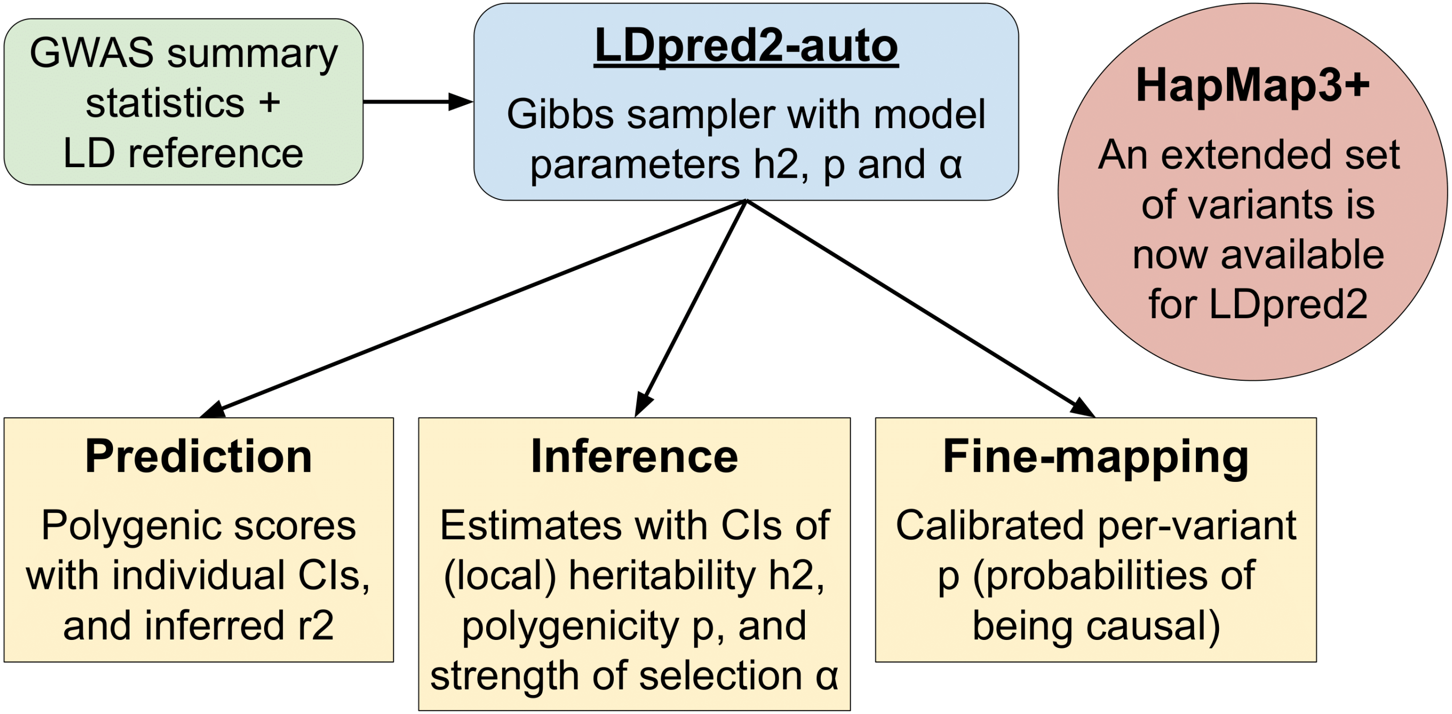
- We have recommended to use a set of 1,054,330 HapMap3 variants
    - good coverage of the genome
    - generally well imputed and available in most studies
- We now provide an extended set with 37% more variants
    - designed to maximize tagging of 11.5M common variants in diverse genetic ancestries
    - called HapMap3+
- Using this new set of variants, in UK Biobank analyses, *on average*,
    - we capture 12% more SNP heritability h^2
    - obtain 6% more predictive performance r^2

⚠ Using more variants won't necessarily give you better polygenic scores.

UKBB results with HapMap3+ (1.4M) vs HapMap3 (1M)



Overview of what LDpred2-auto can now provide



GWAS summary statistics + LD reference

LDpred2-auto

Gibbs sampler with model parameters h2, p and α

HapMap3+

An extended set of variants is now available for LDpred2

Prediction

Polygenic scores with individual CIs, and inferred r2

Inference

Estimates with CIs of (local) heritability h2, polygenicity p, and strength of selection α

Fine-mapping

Calibrated per-variant p (probabilities of being causal)

Future development

- Design automated decisions for choosing parameters such as `use_MLE`
- Provide means for enhanced quality control of GWAS summary statistics
- Extend LDpred2-auto for
  - using more variants
  - incorporating functional annotations
  - multi-ancestry prediction and inference
- and for (smaller priority):
  - using multiple phenotypes and estimating genetic correlation
  - imputing GWAS summary statistics

Acknowledgments

Co-authors:

- Bjarni J. Vilhjálmsson (Aarhus Uni, DK)
- Julyan Arbel (INRIA Grenoble, FR)
- Hugues Aschard (Pasteur Institute, FR)
- Bogdan Pasaniuc (UCLA, CA, USA)
- Yi Ding (UCLA)
- Clara Albiñana (Aarhus Uni)

Funding:

- Danish National Research Foundation
  (Niels Bohr Professorship to Prof. John McGrath)
- Lundbeck Foundation
  (Fellowship to Prof. B.J. Vilhjálmsson)

Thanks!

Presentation available at bit.ly/ldpred2_wcpg2023

privefl

Slides created via the R package **xaringan**