

# MLBF-PRS: A MACHINE LEARNING MODEL DEVELOPMENT AND BENCHMARKING FRAMEWORK FOR POLYGENIC RISK SCORES

**Anonymous authors**

Paper under double-blind review

## ABSTRACT

In contrast to other genomic tasks, the development of machine learning-based individual-level, genome-wide predictive models, typically termed polygenic risk scores (PRS), have shown little improvement from the use of complex machine learning (ML) methods. This disparity can be attributed to challenges in accessibility, comparability across studies, and a lack of development and evaluation guidelines that enable reproducibility. Sequence-based genomic tasks benefit from benchmarks, which have proven to be fruitful in the advancement of machine learning model development across domains.

To overcome the challenges present in the development of ML-based PRS models, we introduce MLBF-PRS, a novel framework as a catalyst to promote and accelerate the development of ML-based solutions. The framework provides flexible Nextflow DSL2 pipelines that enable parallel comparison of ML models (SVMs, random forests, neural networks) against established statistical PRS methods, comprehensive quality control and data preparation modules following PRS-specific best practices, and automated tracking of model parameters, trained weights, and configurations to ensure full reproducibility.

We describe the usage of MLBF-PRS to showcase how this framework provides accessibility, where, in most cases, the setup and evaluation of PRS models can be time-consuming and require navigation of multiple software tools. The standardised and reproducible dataset-specific benchmarking through MLBF-PRS offers a practical alternative to traditional open benchmarks. We make our framework openly available and continue expanding its capabilities.

## 1 INTRODUCTION

Polygenic risk scores (PRS) estimate an individual’s population level risk of disease by accumulating the effect of multiple variations, single-nucleotide polymorphisms (SNPs), across the genome. In most cases these SNPs are substitutions of a single nucleotide (A,C,T,G) on the genome. The classical PRS of an individual  $j$  in a population is defined as the sum over the number of affected alleles of the SNPs of interest multiplied by their respective effect sizes:

$$PRS_j = \sum_{i=1}^N \beta_i d_i, \quad (1)$$

where  $\beta_i \in \mathbb{R}$ ,  $d_i \in \{0, 1, 2\}$  and  $N \in \mathbb{N}_+$  the number of SNPs. The effect sizes in the classical model are derived from genome-wide association studies (GWAS). To this day, a multitude of extensions to the classical model have been developed (Ndong Sima et al., 2024; Benoumhani et al., 2025). Contrary to other genomic data tasks such as variant classification, functional annotation or DNA generation (Cheng et al., 2023; Avsec et al., 2025; Brixi et al., 2025), state-of-the-art models in PRS estimation do not rely on machine learning algorithms. These aforementioned genomic data machine learning models use genome sequence data, which differs quite significantly from the genotype array data used for PRS development, for which the data consists of single measurements of variations at single loci across the whole genome. Although there are ongoing efforts to develop

054 machine learning-based PRS, the adaptation of machine learning in PRS development significantly  
055 lags behind the aforementioned domains.  
056

### 057 1.1 DATA ACCESSIBILITY 058

059 One obstacle to the development of machine learning-based PRS models exists due to data accessi-  
060 bility restrictions, which is a common issue with medical data that directly influences reproducibil-  
061 ity (McDermott et al., 2021; McDermott, 2025). For genomic data problems involving genomic  
062 sequences, this issue can partially be circumvented through the use of reference genomes and open  
063 datasets. However, the evaluation and development of machine learning-based PRS methods in most  
064 cases requires access to individual level data from cohorts (Wang et al., 2023), such as the popular  
065 resource UK Biobank (Bycroft et al., 2018). Access to such informative and rich resources is subject  
066 to individual approval and data is not publicly available due to ethical guidelines, because individ-  
067 ual level data contains sensitive information about study participants. This can slow down model  
068 development and stands in stark contrast to open benchmarking datasets or competitions commonly  
069 found in machine learning model development. At the same time, this data is necessary to improve  
070 the clinical applicability of a PRS. However, development of methodological concepts can proceed  
071 on synthetic datasets (Su et al., 2011; Tang & Liu, 2019; Wharrie et al., 2023), where genotypes and  
072 observable traits, phenotypes, are simulated based on openly available reference genomes such as  
073 the 1000 genomes (Auton et al., 2015). Nevertheless, synthetic data cannot fully substitute real data.  
074 Since data generation is controlled, this prior knowledge about the data can introduce bias in model  
075 development.

### 076 1.2 REQUIREMENTS FOR MACHINE LEARNING MODEL DEVELOPMENT 077

078 A recent survey of advancements in the development of deep learning-based PRS methods high-  
079 lighted difficulties in comparability and reproducibility across phenotypes and datasets (Schuran  
080 et al., 2025). Dedicated deep learning architectures for PRS prediction were generally not evaluated  
081 against another in model development and the majority of publications reported favourable results  
082 towards their own models. A meta analysis of predictive performances is made difficult due to the  
083 evaluation of models on differing datasets in term of population structure, data size, features, and  
084 different outcomes of interest. In the analyses concerning the development and evaluation of deep  
085 learning-based models we can observe conflicting results concerning predictive performance im-  
086 provements with respect to statistical models. Due to a lack of a common benchmark for machine  
087 learning-based models no definitive conclusion on advantages through machine learning models can  
088 be made. We attribute the absence of such a benchmark to difficulties in reproducibility, since there  
089 have been benchmarking efforts for statistical models. These issues can, in most cases, be attributed  
090 to the lack of shared data, code and trained model weights.

091 Due to the underexplored development of machine learning-based PRS models with respect to ge-  
092 nomic sequence-related tasks, we have identified the need for a benchmarking and development  
093 framework to aid the development of models. We identify the following necessary properties for a  
094 framework based on the burden of data accessibility and reproducibility in machine learning model  
095 development:

#### 096 **Flexibility:**

097 The framework needs to be able to operate on a multitude of different datasets, differing in size,  
098 population structure and including phenotypes of quantitative or categorical type. Additionally, the  
099 framework should be easy to use to promote its use in the research community, as well as easily  
100 extendable to new models and customisable for evaluation and data preparation.

#### 101 **Comparability:**

102 To validate newly developed methods rigorously it is necessary to compare their prediction against  
103 state-of-the-art PRS models. Therefore, the framework should integrate benchmarking standards  
104 that allow for the immediate comparison across models, machine learning-based as well as estab-  
105 lished statistical models. Benchmarking standards have shown to be fruitful for the development of  
106 machine learning models in tasks such as image classification (Deng et al., 2009).

#### 107 **Reproducibility:**

108 While the comparison against established PRS models yields insights about predictive performance,  
109 it is equally important to evaluate reproducibility of newly developed models. Reproducibility aids  
110 in validation of models across datasets as well as investigations about causality, and cross-ancestry  
111 transferability.

### 113 1.3 SCOPE OF THIS WORK

114  
115 In this work we propose MLBF-PRS, the first benchmarking and development framework for ma-  
116 chine learning-based PRSs. We make contributions to address the above mentioned requirements.  
117 Firstly, to address flexibility, our framework focuses on ease of usage and extendability through the  
118 use of Nextflow pipelines, which connect core functionalities and processing steps. This modular  
119 approach will promote the establishment of the pipeline as a valuable choice for model development  
120 and benchmarking. The framework offers a new way of benchmarking machine learning models  
121 on restricted datasets due to a stringent end-to-end model training and evaluation that allows for the  
122 immediate comparison of established PRS models to the machine learning model in development.  
123 Crucially, recently developed models often claim superior performance in their original publications  
124 (Jayasinghe et al., 2024). Therefore, to facilitate comparability across datasets, and to allow for  
125 quick validation of studies, all parameters associated with data preparation, model evaluation and  
126 model training are recorded. In this way, the results of an experiment can be reproduced by sharing  
127 configuration files if access to a dataset is granted. Apart from sharing of training configurations,  
128 the framework also promotes the sharing of trained model parameters to validate models without  
129 the need for data sharing, similar to the already established sharing of summary statistics of GWAS  
130 studies (Sollis et al., 2023). An in depth overview of novel contributions and its explicit usage inside  
131 the framework is discussed further in section 2.

## 132 2 THE DEVELOPMENT AND BENCHMARKING FRAMEWORK

133  
134 The framework aims to promote the development of machine learning-based PRS models as an al-  
135 ternative to an open benchmarking dataset. To provide similar benefits as a benchmarking dataset,  
136 we incorporate flexibility, comparability and reproducibility together with PRS specific recommen-  
137 dations to present a state-of-the-art framework that fills the gap for machine learning-based PRS  
138 model development.

### 139 2.1 FLEXIBILITY

140  
141 The framework consists of multiple modules that, when combined together in a single end-to-end  
142 pipeline, allow for the automatic training and evaluation of machine learning models. Each module  
143 consists of multiple scripts that can be customised or expanded to fit an algorithm’s properties or  
144 include new algorithms. As visualised in Fig. 1 the three core modules are related to data prepara-  
145 tion, which is known as quality control in the context of PRS, model training, and model evaluation.  
146 The core module for model training is separated into three different streams of models due to their  
147 specific requirements. Here, established statistical PRS models or tools can be trained or calculated,  
148 while machine learning models can be trained with a separate training pipeline for deep learning  
149 models.

150 One strength of the framework lies in its modularity. To include a new algorithm it is only necessary  
151 to create connection points in the form of scripts to the quality control module for input and the  
152 evaluation model for its output. Apart from adding new algorithms, it is also possible to include  
153 different or additional data preparation or augmentation steps at the connection points of each quality  
154 control step. Lastly, as the prediction results also report raw scores and effect sizes, any desired  
155 metric may be inferred from a new evaluation script.

156 The framework is built on Nextflow (Di Tommaso et al., 2017), a software tool that allows for the  
157 orchestration and execution of scripts as pipelines. Its advantage lies in the ability to run independent  
158 scripts in parallel while sequential steps wait for output from dependent scripts. This effectively  
159 allows for the execution of the three streams in the core module in parallel and for the training of  
160 models in parallel. However, this capability is restricted by available resources. To orchestrate  
161 parallel computing, Nextflow allows for the allocation of resources to a specific pipeline through  
configuration files. The statistical PRS tools implemented in this framework are all callable via

162 command line interface, which makes them easily implementable into a Nextflow pipeline. Python  
163 scripts and R scripts are all created as command line tools to make them implementable and capable  
164 for rapid testing.

165 The framework utilizes the PLINK file format (Purcell et al., 2007) as well as the PLINK software  
166 for quality control. We specifically choose to work with the PLINK file format as rich data sources  
167 like the UKB or All of Us (Bick et al., 2024) release genetic data in the PLINK file format. PLINK  
168 itself also offers commands to convert other file formats like .vcf into PLINK files. We make the  
169 data accessible for use in Python-based machine learning frameworks scikit-learn (Pedregosa et al.,  
170 2011) and Pytorch (Paszke et al., 2019) through specialised data loading modules.

## 171 172 173 2.2 COMPARABILITY

174  
175  
176 To facilitate comparability like an open benchmark, the framework utilizes the development and  
177 evaluation pipeline in such a way that, independent of the dataset, a comprehensive benchmark  
178 can be created. A dataset specific benchmark acknowledges the unique challenges encountered  
179 when developing phenotype-specific models. While the environment surrounding the PRS model  
180 development can be transferred between datasets (i.e. quality control, data loading, or statistical  
181 analysis) the phenotype specific modelling can differ significantly, because phenotypes differ in  
182 heritability, gene interactions or biological pathways.

183 Consequently, the evaluation and development framework offers an environment that follows recom-  
184 mendations for PRS-specific model evaluation. The data preparation module includes an extensive  
185 quality control proposed by Choi et al. (2020). The quality control is applied to summary statistics  
186 as well as the individual level data. Summary statistics are typically not used for the development  
187 of machine learning-based methods, but are still necessary for the statistical models considered for  
188 comparison. Apart from filtering SNPs and individuals by genotyping rate, sample missingness,  
189 Hardy-Weinberg equilibrium, heterozygosity, minor allele frequency and imputation score through  
190 the use of PLINK, the quality control pipeline includes the following additional steps: removal of  
191 ambiguous SNPs, resolving or removal of non-resolvable mismatching SNPs, removal of duplicate  
192 SNPs, removal of individuals with mismatching reported and genetic sex and lastly the removal of  
193 closely related individuals. Besides the quality control steps the data preparation modules allows for  
194 the calculation of principal components, which are commonly used in many PRS models to account  
195 for confounding effects through population structures in datasets. Lastly filtering of individuals,  
196 SNPs and data augmentation of covariates is also done in the first module of the framework. In con-  
197 trast to Pain et al. (2021), we do not restrict the choice of SNPs to HapMap3, we intentionally leave  
198 this choice open due to flexible choices for machine learning-based models, but report on SNPs at  
the end of the pipeline.

199 Between the data preparation and model training lies an integral part of comparable model develop-  
200 ment which is data splitting into a training dataset and a hold-out dataset for later model evaluation.  
201 Similar to Pain et al. (2021) we integrate k-fold cross-validation.

202 Due to the majority of state-of-the-art PRS model development focusing on statistical models it is  
203 essential to compare any machine learning-based model against these established methods. There-  
204 fore, we integrate a representative collection of such methods to benchmark against the machine  
205 learning-based models. We chose to integrate PRSice-2 (Choi & O'Reilly, 2019), to represent the  
206 gold standard approach of clumping and thresholding (CT), an extension to CT called stacked CT  
207 Privé et al. (2019), LDpred2 Privé et al. (2021), Lassosum Mak et al. (2017) and Lassosum2 Privé  
208 et al. (2022), PRS-CS Ge et al. (2019) and PRS-CSx Ruan et al. (2022), SBayesRC Zheng et al.  
209 (2024) and PRSet Choi et al. (2023) to capture a variety from the landscape of current state-of-the-  
210 art solutions. Some of the models require additional data such as LD reference panels, pathway  
211 information or ancestry specific data, which is made available in the pipeline.

212 In contrast to the statistical models, the machine learning-based models are in most cases trained in  
213 a supervised setting to predict a phenotype. In the case of a binary phenotype, such as cancer status,  
214 the class likelihood is used as a PRS. In that case genotype data is used together with covariates for  
215 training. The data loader combines and provides this data for every individual during model training  
or evaluation. The covariates need to be specified otherwise all of them are used.

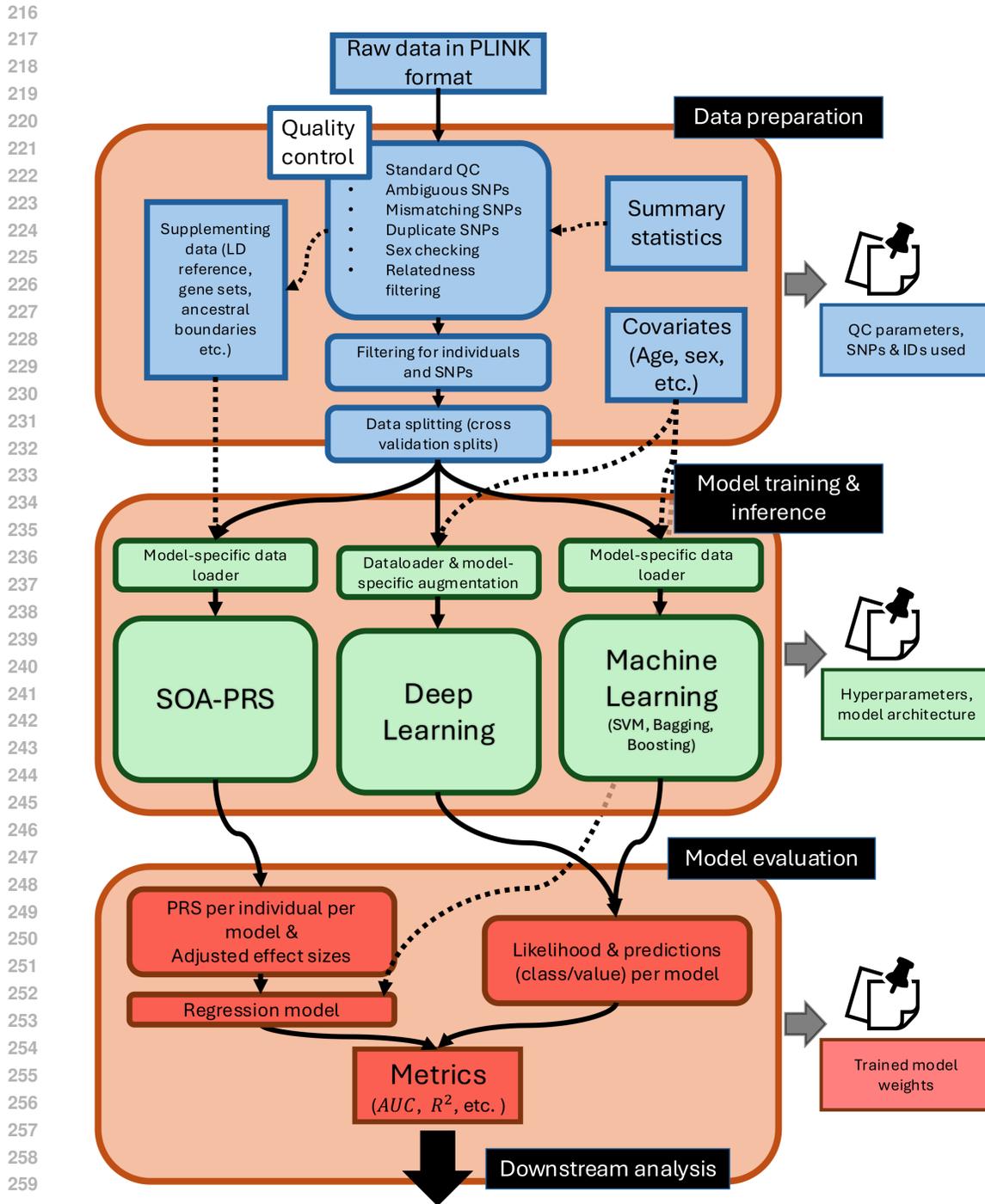


Figure 1: Visualisation of the framework, divided by the three core modules. Each module retains data to facilitate reproducibility.

To create a predictive model from the statistical models, typically, a linear, for continuous phenotypes, or a logistic, for binary phenotypes, regression model is fitted on the PRS together with covariates. Therefore, we report the PRS for each of those models per individual as well as the effect size for each SNP per model. With this raw data it is possible to then fit predictive models per method or even create ensemble models. From here metrics can be calculated to compare perfor-

mance across models. We recommend following reporting standards as proposed by (Wand et al., 2021). These include metrics such as  $R^2$ , AUC, reclassification indices but also statistical analyses like distributions of cases and controls.

In summary, the framework offers a standardised and concise end-to-end benchmarking approach for the development of new PRS models or inference of a selection of models which can be easily customised and extended.

### 2.3 REPRODUCIBILITY

An equally important problem in the development of PRS models lies in the validation of a model on external data which it has not seen during model training. To achieve such reproducibility the framework offers essentially two capabilities that aid the validation of models. Firstly, every parameter in the pipeline, from quality control, over model training, to evaluation, is specified in configuration files associated with each of the three modules, as shown in Fig 1. In this way, the whole pipeline or single steps should be reproducible if access to the same data resource is given. For instance, in favourable circumstances, where a dataset is openly accessible, e.g. available on dbGaP (Tryka et al., 2014), a validation study would only need to download the dataset and with shared configuration files an entire analysis could be easily repeated. A comprehensive overview of PRS development and calculation through these configuration files also promotes the clinical transferability of PRSs (Xiang et al., 2024). However, as previously mentioned, access to datasets might be restricted. In such cases we recommend opting for additional sharing of model architecture and trained weights in case of machine learning model development. The framework, by default, saves trained weights and each model definition is stored in separate files. Sharing a model in this way retains information in its trained weights and sharing of original data is not necessary, similar to sharing results of GWAS analyses in form of summary statistics or published PRSs on the PGS catalog (Lambert et al., 2021). Apart from the sharing of trained machine learning model weights, the modified effect sizes of each statistical model are saved as well. The sharing of learned machine learning model weights together with finetuning pretrained models remains an underexplored area in the field of PRSs and our framework makes it more accessible.

## 3 USAGE OF THE FRAMEWORK

As MLPBF-PRS is a modular framework it can be accessed on multiple levels from top to bottom. At the highest level a whole analysis across all core modules can be conducted by running the script `run_pipelines.sh`. This script then runs multiple Nextflow pipelines following the illustration of information flow shown in Fig. 1. Each pipeline can however be executed independently given its input data exists. As an example the model training and inference pipelines require input data generated from the quality control module. Pipelines are called through Nextflow as shown below in the case of the statistical PRS models or for deep learning-based PRS approaches:

```
nextflow run workflows/prs_models_pipeline.nf
           --params-file workflows/config/params_prs.yaml
nextflow run workflows/dl_prs.nf
           --params-file workflows/config/dl_config.yaml
```

Inside the `.yaml`-files model parameters are set to promote reproducibility. At the foundation of the framework lie Python scripts, R scripts and command line tools. The scripts or tools are embedded into Nextflow pipeline, but can also accessed for rapid testing. As an example the training script for deep learning models is called inside a Nextflow pipeline as shown below. The `params.xxx` inputs in this case are defined inside a `.yaml` file while another configuration file is passed to the training routine via `${params_file}`.

```
python ${params.base_dir}/bin/train_model.py \
      --genotype_file ${genotype_data} \
      --phenotype_file ${phenotypes} \
      --indices_file ${indices} \
      --params_file ${params_file} \
      --fold ${fold} \
```

```

324         --max_epochs ${params.max_epochs} \
325         --batch_size ${params.batch_size} \
326         --output_model model_fold_${fold}.pt \
327         --output_metrics metrics_fold_${fold}.json \
328         --output_log training_log_fold_${fold}.csv \
329         ${wandb_args}
330

```

The multi-levelled approach to the framework enables in depth as well as high level access to the field of PRS development.

## 4 DISCUSSION

### 4.1 RELATED RESEARCH

There exist several open benchmarks across genomic data problems that have successfully produced novel machine learning models. The precisionFDA Truth Challenge 2.0 Olson et al. (2022), which utilizes the Genome in a Bottle benchmark, motivated the development of new machine learning approaches to improve variant calling. CAGI (Jain et al., 2024) and CAFA (Zhou et al., 2019) aim to promote model development for prediction of genetic variant impact and protein function prediction respectively. In CAGI challenges machine learning models have shown to outperform statistical models Savojardo et al. (2019) and in CAFA3 a machine learning ensemble model performed best (Zhou et al., 2019). Recent efforts have evaluated genomic foundation models systematically. Patel et al. (2024) introduce a DNA language model (DNALM) benchmark for short, local sequences evaluating models like HyeanaDNA, DNABERT or Nucleotide Transformer. Similarly, Kao et al. (2024) evaluated DNALMs for long range dependencies. Liu et al. (2024) propose a benchmarking framework for DNA sequences focusing on modularity and reproducibility like MLBF-PRS. Notably, these models are all trained on openly available sequence data. Kopp et al. (2020) offer a flexible Python module, named Janguu, for sequence data management to train neural networks.

However, PRSs present unique challenges: integrating genome-wide signals rather than local sequences, predicting clinical outcomes rather than molecular phenotypes, and dealing with individual level data. While there is no shortage of development of machine learning-based PRSs there exist no comprehensive open benchmarks across such models. Benchmarks mostly focus on statistical models or treat machine learning models rather generically (Zhao et al., 2024; Wang et al., 2023; Allegrini et al., 2022), even though there exist a multitude of specifically designed neural network architectures (Schuran et al., 2025). These newly developed neural networks benchmark mainly against statistical models and not other task-specific neural networks, although the majority of studies report performance improvements of their respective novel method. For statistical models efforts have been made to make PRS models more accessible through pipelines that integrate multiple models. PGS-Depot focuses on a pipeline for summary statistic based PRSs (Cao et al., 2024). Yang & Zhou (2022), Lee et al. (2023) and Jin et al. (2025) provide a web based solution for polygenic risk prediction, addressing accessibility and ease of usage. Yaraş et al. (2025) and Pain et al. (2024) focus on a standardised and reproducible pipeline, where PGSEXplorer also uses Nextflow pipelines. Pham et al. (2022) address diverse populations in their standardised benchmarking software. None of these existing pipelines investigates machine learning-based solutions.

Witteveen et al. (2022) propose a different approach in creating an open benchmark. Here, a synthetic dataset based on real genotypes is generated that yields privacy preserving open dataset. They show that absolute and relative performance of models can be accurately evaluated using statistical PRS models. Mendes et al. (2025) suggest that utilizing synthetic data can accelerate machine learning model development even for rare diseases.

### 4.2 MLBF-PRS TO BRIDGE THE GAP BETWEEN STATISTICAL AND MACHINE LEARNING-BASED PRS

We position MLBF-PRS in the existing gap between established benchmarking approaches from the machine learning community for DNALMs and the computational tools such as pipelines and frameworks from the bioinformatics domain. The focus of MLBF-PRS is utilizing the accessibility of a benchmarking pipeline to catalyse reproducible machine learning-model development. We

378 suspect current PRS benchmarks and frameworks ignore machine learning models due to a lack  
379 of reproducibility, therefore, our framework encourages the development of reproducible research.  
380 In this way provide a platform for machine learning-based models with the aim to benchmark them  
381 against another, while still providing statistical model predictions as a baseline.

#### 382 383 4.3 LIMITATIONS AND FUTURE DEVELOPMENTS 384

385 Many analyses of PRS include the investigation of interpretability to find or validate known causal  
386 biological pathways. Explainable artificial intelligence has the potential to yield such insights from  
387 machine learning models (Novakovsky et al., 2023), therefore, we plan to integrate methods such  
388 as DeepLift (Shrikumar et al., 2017) or Grad-CAM (Selvaraju et al., 2020) into our framework. So  
389 far during development we tested our framework solely on synthetic data generated from a refer-  
390 ence panel that used whole genome sequencing with high coverage (Byrska-Bishop et al., 2022).  
391 When dealing with real datasets, it is common to have to account for missing data, imputations,  
392 possible confounding factors and very imbalanced cases and controls. In the future we plan to ap-  
393 ply the framework to large biobank data such as the UKB to test robustness of the pipeline and to  
394 further expand its capabilities. Perelygin et al. (2025) have shown superior improved of non-linear  
395 machine learning models on simulated epistasis phenotypes with respect to linear models, we plan  
396 to validate performance gains of non-linear models on real data in the future with MLBF-PRS. So  
397 far the framework lacks its own data generation pipeline, we think, that the privacy preserving data  
398 generation proposed by Witteveen et al. (2022) could further improve the framework. Consequently  
399 we plan to integrate automatic data generation functionality to further improve accessibility. Apart  
400 from that, we only integrate a collection of all the statistical models available, but the flexibility and  
401 modularity of the framework allows for the integration of additional models in the future. Lastly,  
402 we suggest that the reproducible pipeline could be used to enable privacy preserving model training  
403 through federated learning (Sabha, 2025).

## 404 5 CONCLUSION 405

406 Adoption of machine learning-based solutions for PRSs lags behind their statistical counterparts as  
407 well as the establishment of machine learning-based models as state-of-the-art models in other ge-  
408 netic data problems. Here, we identified three key factors that we believe hinder progress in machine  
409 learning-based PRS development, namely accessibility issues to valuable datasets, compromised  
410 comparability due to the diverse nature of datasets and the lack of an open benchmarking datasets,  
411 common in deep learning-model development, and finally a lack in reproducibility. Therefore, we  
412 introduce MLBF-PRS, a framework designed to make the development and inference of machine  
413 learning-based PRS models more accessible while providing an alternative solution to open bench-  
414 marking datasets. The framework acknowledges the diversity of existing data and aims to bridge the  
415 current gap in realizing the modeling advantages that machine learning approaches offer for SNP  
416 data analysis. MLBF-PRS is designed with extensibility in mind, making it easy for the research  
417 community to contribute enhancements and adopt the software to accelerate model development.  
418 Given that machine learning models have demonstrated promising results in cross-ancestry appli-  
419 cations (Elgart et al., 2022) and variant discovery (Lakiotaki et al., 2023), we believe that machine  
420 learning solutions for PRS warrant further investigation and validation. MLBF-PRS serves as a  
421 practical tool to advance progress in this rapidly evolving field.

#### 422 LLM USAGE DECLARATION 423

424 Large Language Models have been used for editorial polishing (spelling and grammar checks) and  
425 as a coding assistant to generate and debug code.

## 427 REFERENCES 428

429 Andrea G. Allegrini, Jessie R. Baldwin, Wikus Barkhuizen, and Jean-Baptiste Pingault. Research  
430 review: A guide to computing and implementing polygenic scores in developmental research.  
431 *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 63(10):1111–1124, October  
2022. ISSN 0021-9630. doi: 10.1111/jcpp.13611.

432 Adam Auton, Gonalo R. Abecasis, David M. Altshuler, Richard M. Durbin, Gonalo R. Abeca-  
433 sis, David R. Bentley, Aravinda Chakravarti, Andrew G. Clark, Peter Donnelly, Evan E. Eichler,  
434 Paul Flicek, Stacey B. Gabriel, Richard A. Gibbs, Eric D. Green, Matthew E. Hurles, Bartha M.  
435 Knoppers, Jan O. Korbel, Eric S. Lander, Charles Lee, Hans Lehrach, Elaine R. Mardis, Gab-  
436 bor T. Marth, Gil A. McVean, Deborah A. Nickerson, Jeanette P. Schmidt, Stephen T. Sherry, Jun  
437 Wang, Richard K. Wilson, Richard A. Gibbs, Eric Boerwinkle, Harsha Doddapaneni, Yi Han,  
438 Viktoriya Korchina, Christie Kovar, Sandra Lee, Donna Muzny, Jeffrey G. Reid, Yiming Zhu, Jun  
439 Wang, Yuqi Chang, Qiang Feng, Xiaodong Fang, Xiaosen Guo, Min Jian, Hui Jiang, Xin Jin,  
440 Tianming Lan, Guoqing Li, Jingxiang Li, Yingrui Li, Shengmao Liu, Xiao Liu, Yao Lu, Xuedi  
441 Ma, Meifang Tang, Bo Wang, Guangbiao Wang, Honglong Wu, Renhua Wu, Xun Xu, Ye Yin,  
442 Dandan Zhang, Wenwei Zhang, Jiao Zhao, Meiru Zhao, Xiaole Zheng, Eric S. Lander, David M.  
443 Altshuler, Stacey B. Gabriel, Namrata Gupta, Neda Gharani, Lorraine H. Toji, Norman P. Gerry,  
444 Alissa M. Resch, Paul Flicek, Jonathan Barker, Laura Clarke, Laurent Gil, Sarah E. Hunt, Gavin  
445 Kelman, Eugene Kulesha, Rasko Leinonen, William M. McLaren, Rajesh Radhakrishnan, Asier  
446 Roa, Dmitriy Smirnov, Richard E. Smith, Ian Streeter, Anja Thormann, Iliana Toneva, Brendan  
447 Vaughan, Xiangqun Zheng-Bradley, David R. Bentley, Russell Grocock, Sean Humphray, Terena  
448 James, Zoya Kingsbury, Hans Lehrach, Ralf Sudbrak, Marcus W. Albrecht, Vyacheslav S. Am-  
449 stislavskiy, Tatiana A. Borodina, Matthias Lienhard, Florian Mertes, Marc Sultan, Bernd Timmer-  
450 mann, Marie-Laure Yaspo, Elaine R. Mardis, Richard K. Wilson, Lucinda Fulton, Robert Fulton,  
451 Stephen T. Sherry, Victor Ananiev, Zinaida Belaia, Dimitriy Beloslyudtsev, Nathan Bouk, Chao  
452 Chen, Deanna Church, Robert Cohen, Charles Cook, John Garner, Timothy Hefferon, Mikhail  
453 Kimelman, Chunlei Liu, John Lopez, Peter Meric, Chris O’Sullivan, Yuri Ostapchuk, Lon Phan,  
454 Sergiy Ponomarov, Valerie Schneider, Eugene Shekhtman, Karl Sirotkin, Douglas Slotta, Hua  
455 Zhang, Gil A. McVean, Richard M. Durbin, Senduran Balasubramaniam, John Burton, Petr  
456 Danecek, Thomas M. Keane, Anja Kolb-Kokocinski, Shane McCarthy, James Stalker, Michael  
457 Quail, Jeanette P. Schmidt, Christopher J. Davies, Jeremy Gollub, Teresa Webster, Brant Wong,  
458 Yiping Zhan, Adam Auton, Christopher L. Campbell, Yu Kong, Anthony Marcketta, Richard A.  
459 Gibbs, Fuli Yu, Lilian Antunes, Matthew Bainbridge, Donna Muzny, Aniko Sabo, Zhuoyi Huang,  
460 Jun Wang, Lachlan J. M. Coin, Lin Fang, Xiaosen Guo, Xin Jin, Guoqing Li, Qibin Li, Yingrui  
461 Li, Zhenyu Li, Haoxiang Lin, Binghang Liu, Ruibang Luo, Haojing Shao, Yinlong Xie, Chen Ye,  
462 Chang Yu, Fan Zhang, Hancheng Zheng, Hongmei Zhu, Can Alkan, Elif Dal, Fatma Kahveci, Gab-  
463 bor T. Marth, Erik P. Garrison, Deniz Kural, Wan-Ping Lee, Wen Fung Leong, Michael Stromberg,  
464 Alistair N. Ward, Jiantao Wu, Mengyao Zhang, Mark J. Daly, Mark A. DePristo, Robert E. Hand-  
465 saker, David M. Altshuler, Eric Banks, Gaurav Bhatia, Guillermo del Angel, Stacey B. Gabriel,  
466 Giulio Genovese, Namrata Gupta, Heng Li, Seva Kashin, Eric S. Lander, Steven A. McCarroll,  
467 James C. Nemes, Ryan E. Poplin, Seungtae C. Yoon, Jayon Lihm, Vladimir Makarov, Andrew G.  
468 Clark, Srikanth Gottipati, Alon Keinan, Juan L. Rodriguez-Flores, Jan O. Korbel, Tobias Rausch,  
469 Markus H. Fritz, Adrian M. Stutz, Paul Flicek, Kathryn Beal, Laura Clarke, Avik Datta, Javier  
470 Herrero, William M. McLaren, Graham R. S. Ritchie, Richard E. Smith, Daniel Zerbino, Xi-  
471 angqun Zheng-Bradley, Pardis C. Sabeti, Ilya Shlyakhter, Stephen F. Schaffner, Joseph Vitti,  
472 David N. Cooper, Edward V. Ball, Peter D. Stenson, David R. Bentley, Bret Barnes, Markus  
473 Bauer, R. Keira Cheetham, Anthony Cox, Michael Eberle, Sean Humphray, Scott Kahn, Lisa  
474 Murray, John Peden, Richard Shaw, Eimear E. Kenny, Mark A. Batzer, Miriam K. Konkel, Jeri-  
475 lyn A. Walker, Daniel G. MacArthur, Monkol Lek, Ralf Sudbrak, Vyacheslav S. Amstislavskiy,  
476 Ralf Herwig, Elaine R. Mardis, Li Ding, Daniel C. Koboldt, David Larson, Kai Ye, Simon Gravel,  
477 The 1000 Genomes Project Consortium, Corresponding authors, Steering committee, Production  
478 group, Baylor College of Medicine, BGI-Shenzhen, Broad Institute of MIT, Harvard, Coriell Insti-  
479 tute for Medical Research, European Bioinformatics Institute European Molecular Biology Labo-  
480 ratory, Illumina, Max Planck Institute for Molecular Genetics, McDonnell Genome Institute  
481 at Washington University, US National Institutes of Health, University of Oxford, Wellcome  
482 Trust Sanger Institute, Analysis group, Affymetrix, Albert Einstein College of Medicine, Bilkent  
483 University, Boston College, Cold Spring Harbor Laboratory, Cornell University, European Molec-  
484 ular Biology Laboratory, Harvard University, Human Gene Mutation Database, Icahn School  
485 of Medicine at Mount Sinai, Louisiana State University, Massachusetts General Hospital, McGill  
University, and NIH National Eye Institute. A global reference for human genetic variation. *Nature*, 526(7571):68–74, October 2015. ISSN 1476-4687. doi: 10.1038/nature15393.

Žiga Avsec, Natasha Latysheva, Jun Cheng, Guido Novati, Kyle R. Taylor, Tom Ward, Clare By-  
croft, Lauren Nicolaisen, Eirini Arvaniti, Joshua Pan, Raina Thomas, Vincent Dutordoir, Matteo

- 486 Perino, Soham De, Alexander Karollus, Adam Gayoso, Toby Sargeant, Anne Mottram, Lai Hong  
487 Wong, Pavol Drotár, Adam Kosiorek, Andrew Senior, Richard Tanburn, Taylor Applebaum,  
488 Souradeep Basu, Demis Hassabis, and Pushmeet Kohli. Alphagenome: advancing regulatory vari-  
489 ant effect prediction with a unified dna sequence model. pp. 2025.06.25.661532, June 2025. doi:  
490 10.1101/2025.06.25.661532. URL [https://www.biorxiv.org/content/10.1101/  
491 2025.06.25.661532v1](https://www.biorxiv.org/content/10.1101/2025.06.25.661532v1).
- 492 Sara Benoumhani, Areej Al-Wabil, Niddal Imam, Bashayer Alfawaz, Amaan Zubairi, Dalal Aldos-  
493 sary, and Mariam AlEissa. A review of methods and software for polygenic risk score analysis.  
494 *PeerJ Computer Science*, 11:e3039, August 2025. ISSN 2376-5992. doi: 10.7717/peerj-cs.3039.  
495
- 496 Alexander G. Bick, Ginger A. Metcalf, Kelsey R. Mayo, Lee Lichtenstein, Shimon Rura, Robert J.  
497 Carroll, Anjene Musick, Jodell E. Linder, I. King Jordan, Shashwat Deepali Nagar, Shivam  
498 Sharma, Robert Meller, Melissa Basford, Eric Boerwinkle, Mine S. Cicek, Kimberly F. Do-  
499 heny, Evan E. Eichler, Stacey Gabriel, Richard A. Gibbs, David Glazer, Paul A. Harris, Gail P.  
500 Jarvik, Anthony Philippakis, Heidi L. Rehm, Dan M. Roden, Stephen N. Thibodeau, Scott Top-  
501 per, Ashley L. Blegen, Samantha J. Wirkus, Victoria A. Wagner, Jeffrey G. Meyer, Mine S. Ci-  
502 cek, Donna M. Muzny, Eric Venner, Michelle Z. Mawhinney, Sean M. L. Griffith, Elvin Hsu,  
503 Hua Ling, Marcia K. Adams, Kimberly Walker, Jianhong Hu, Harsha Doddapaneni, Christie L.  
504 Kovar, Mullai Murugan, Shannon Dugan, Ziad Khan, Eric Boerwinkle, Niall J. Lennon, Christina  
505 Austin-Tse, Eric Banks, Michael Gatzen, Namrata Gupta, Emma Henricks, Katie Larsson,  
506 Sheli McDonough, Steven M. Harrison, Christopher Kachulis, Matthew S. Lebo, Cynthia L.  
507 Neben, Marcie Steeves, Alicia Y. Zhou, Joshua D. Smith, Christian D. Frazar, Colleen P. Davis,  
508 Karynne E. Patterson, Marsha M. Wheeler, Sean McGee, Christina M. Lockwood, Brian H.  
509 Shirts, Colin C. Pritchard, Mitzi L. Murray, Valeria Vasta, Dru Leistritz, Matthew A. Richardson,  
510 Jillian G. Buchan, Aparna Radhakrishnan, Niklas Krumm, Brenna W. Ehmen, Sophie Schwartz,  
511 M. Morgan T. Aster, Kristian Cibulskis, Andrea Haessly, Rebecca Asch, Aurora Cremer, Kylee  
512 Degatano, Akum Shergill, Laura D. Gauthier, Samuel K. Lee, Aaron Hatcher, George B. Grant,  
513 Genevieve R. Brandt, Miguel Covarrubias, Eric Banks, Ashley Able, Ashley E. Green, Robert J.  
514 Carroll, Jennifer Zhang, Henry R. Condon, Yuanyuan Wang, Moira K. Dillon, C. H. Albach,  
515 Wail Baalawi, Seung Hoan Choi, Xin Wang, Elisabeth A. Rosenthal, Andrea H. Ramirez, Sokny  
516 Lim, Siddhartha Nambiar, Bradley Ozenberger, Anastasia L. Wise, Chris Lunt, Geoffrey S. Gins-  
517 burg, and Joshua C. Denny. Genomic data in the all of us research program. *Nature*, 627(8003):  
340–346, March 2024. ISSN 1476-4687. doi: 10.1038/s41586-023-06957-x.
- 518 Garyk Brixi, Matthew G. Durrant, Jerome Ku, Michael Poli, Greg Brockman, Daniel Chang,  
519 Gabriel A. Gonzalez, Samuel H. King, David B. Li, Aditi T. Merchant, Mohsen Naghipour-  
520 far, Eric Nguyen, Chiara Ricci-Tam, David W. Romero, Gwanggyu Sun, Ali Taghibakshi, An-  
521 ton Vorontsov, Brandon Yang, Myra Deng, Liv Gorton, Nam Nguyen, Nicholas K. Wang,  
522 Etowah Adams, Stephen A. Baccus, Steven Dillmann, Stefano Ermon, Daniel Guo, Rajesh  
523 Ilango, Ken Janik, Amy X. Lu, Reshma Mehta, Mohammad R. K. Mofrad, Madelena Y. Ng,  
524 Jaspreet Pannu, Christopher Ré, Jonathan C. Schmok, John St John, Jeremy Sullivan, Kevin Zhu,  
525 Greg Zynda, Daniel Balsam, Patrick Collison, Anthony B. Costa, Tina Hernandez-Boussard,  
526 Eric Ho, Ming-Yu Liu, Thomas McGrath, Kimberly Powell, Dave P. Burke, Hani Goodarzi,  
527 Patrick D. Hsu, and Brian L. Hie. Genome modeling and design across all domains of life  
528 with evo 2. pp. 2025.02.18.638918, February 2025. doi: 10.1101/2025.02.18.638918. URL  
529 <https://www.biorxiv.org/content/10.1101/2025.02.18.638918v1>.
- 530 Clare Bycroft, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T. Elliott, Kevin Sharp, Allan  
531 Motyer, Damjan Vukcevic, Olivier Delaneau, Jared O’Connell, Adrian Cortes, Samantha Welsh,  
532 Alan Young, Mark Effingham, Gil McVean, Stephen Leslie, Naomi Allen, Peter Donnelly, and  
533 Jonathan Marchini. The uk biobank resource with deep phenotyping and genomic data. *Nature*,  
534 562(7726):203–209, October 2018. ISSN 1476-4687. doi: 10.1038/s41586-018-0579-z.
- 535 Marta Byrska-Bishop, Uday S. Evani, Xuefang Zhao, Anna O. Basile, Haley J. Abel, Allison A.  
536 Regier, André Corvelo, Wayne E. Clarke, Rajeewa Musunuri, Kshithija Nagulapalli, Susan Fair-  
537 ley, Alexi Runnels, Lara Winterkorn, Ernesto Lowy, Evan E. Eichler, Jan O. Korb, Charles Lee,  
538 Tobias Marschall, Scott E. Devine, William T. Harvey, Weichen Zhou, Ryan E. Mills, Tobias  
539 Rausch, Sushant Kumar, Can Alkan, Fereydoun Hormozdiari, Zechen Chong, Yu Chen, Xiaofei  
Yang, Jiadong Lin, Mark B. Gerstein, Ye Kai, Qihui Zhu, Feyza Yilmaz, Chunlin Xiao, Paul

- 540 Flicek, Soren Germer, Harrison Brand, Ira M. Hall, Michael E. Talkowski, Giuseppe Narzisi,  
541 and Michael C. Zody. High-coverage whole-genome sequencing of the expanded 1000 genomes  
542 project cohort including 602 trios. *Cell*, 185(18):3426–3440.e19, 2022. ISSN 0092-8674, 1097-  
543 4172. doi: 10.1016/j.cell.2022.08.004.
- 544 Chen Cao, Shuting Zhang, Jianhua Wang, Min Tian, Xiaolong Ji, Dandan Huang, Sheng Yang, and  
545 Ning Gu. Pgs-depot: a comprehensive resource for polygenic scores constructed by summary  
546 statistics based methods. *Nucleic Acids Research*, 52(D1):D963–D971, January 2024. ISSN  
547 0305-1048. doi: 10.1093/nar/gkad1029.
- 548 Jun Cheng, Guido Novati, Joshua Pan, Clare Bycroft, Akvilė Žemgulytė, Taylor Applebaum,  
549 Alexander Pritzel, Lai Hong Wong, Michal Zielinski, Tobias Sargeant, Rosalia G. Schneider,  
550 Andrew W. Senior, John Jumper, Demis Hassabis, Pushmeet Kohli, and Žiga Avsec. Accu-  
551 rate proteome-wide missense variant effect prediction with alphamissense. *Science*, 381(6664):  
552 eadg7492, September 2023. doi: 10.1126/science.adg7492.
- 553 Shing Wan Choi and Paul F O’Reilly. Prsice-2: Polygenic risk score software for biobank-scale  
554 data. *GigaScience*, 8(7):giz082, 2019. ISSN 2047-217X. doi: 10.1093/gigascience/giz082.
- 555 Shing Wan Choi, Timothy Shin-Heng Mak, and Paul F. O’Reilly. Tutorial: a guide to performing  
556 polygenic risk score analyses. *Nature Protocols*, 15(99):2759–2772, September 2020. ISSN  
557 1750-2799. doi: 10.1038/s41596-020-0353-1.
- 558 Shing Wan Choi, Judit García-González, Yunfeng Ruan, Hei Man Wu, Christian Porras, Jessica  
559 Johnson, Bipolar Disorder Working group of the Psychiatric Genomics Consortium, Clive J.  
560 Hoggart, and Paul F. O’Reilly. Prset: Pathway-based polygenic risk score analyses and soft-  
561 ware. *PLOS Genetics*, 19(2):e1010624, February 2023. ISSN 1553-7404. doi: 10.1371/journal.  
562 pgen.1010624.
- 563 Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hier-  
564 archical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*,  
565 pp. 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.
- 566 Paolo Di Tommaso, Maria Chatzou, Evan W. Floden, Pablo Prieto Barja, Emilio Palumbo, and  
567 Cedric Notredame. Nextflow enables reproducible computational workflows. *Nature Biotechnol-  
568 ogy*, 35(4):316–319, April 2017. ISSN 1546-1696. doi: 10.1038/nbt.3820.
- 569 Michael Elgart, Genevieve Lyons, Santiago Romero-Brufau, Nuzulul Kurniansyah, Jennifer A.  
570 Brody, Xiuqing Guo, Henry J. Lin, Laura Raffield, Yan Gao, Han Chen, Paul de Vries, Donald M.  
571 Lloyd-Jones, Leslie A. Lange, Gina M. Peloso, Myriam Fornage, Jerome I. Rotter, Stephen S.  
572 Rich, Alanna C. Morrison, Bruce M. Psaty, Daniel Levy, Susan Redline, Paul de Vries, and  
573 Tamar Sofer. Non-linear machine learning models incorporating snps and prs improve polygenic  
574 prediction in diverse human populations. *Communications Biology*, 5(1):856, August 2022. ISSN  
575 2399-3642. doi: 10.1038/s42003-022-03812-z.
- 576 Tian Ge, Chia-Yen Chen, Yang Ni, Yen-Chen Anne Feng, and Jordan W. Smoller. Polygenic pre-  
577 diction via bayesian regression and continuous shrinkage priors. *Nature Communications*, 10(1):  
578 1776, April 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-09718-5.
- 579 Shantanu Jain, Constantina Bakolitsa, Steven E. Brenner, Predrag Radivojac, John Moulton, Susanna  
580 Repo, Roger A. Hoskins, Gaia Andreoletti, Daniel Barsky, Ajithavalli Chellapan, Hoyin Chu,  
581 Navya Dabhiru, Naveen K. Kollipara, Melissa Ly, Andrew J. Neumann, Lipika R. Pal, Eric Odell,  
582 Gaurav Pandey, Robin C. Peters-Petrulewicz, Rajgopal Srinivasan, Stephen F. Yee, Sri Jyothsna  
583 Yeleswarapu, Maya Zuhl, Ogun Adebali, Ayoti Patra, Michael A. Beer, Raghavendra Hosur, Jian  
584 Peng, Brady M. Bernard, Michael Berry, Shengcheng Dong, Alan P. Boyle, Aashish Adhikari,  
585 Jingqi Chen, Zhiqiang Hu, Robert Wang, Yaqiong Wang, Maximilian Miller, Yanran Wang, Yana  
586 Bromberg, Paola Turina, Emidio Capriotti, James J. Han, Kivilcim Ozturk, Hannah Carter, Giu-  
587 lia Babbli, Samuele Bovo, Pietro Di Lena, Pier Luigi Martelli, Castrense Savojardo, Rita Casa-  
588 dio, Melissa S. Cline, Greet De Baets, Sandra Bonache, Orland Díez, Sara Gutiérrez-Enríquez,  
589 Alejandro Fernández, Gemma Montalbán, Lars Ootes, Selen Özkan, Natàlia Padilla, Casandra  
590 Riera, Xavier De la Cruz, Mark Diekhans, Peter J. Huwe, Qiong Wei, Qifang Xu, Roland L.  
591 Dunbrack, Valer Gotea, Laura Elnitski, Gennady Margolin, Piero Fariselli, Ivan V. Kulakovskiy,  
592  
593

- 594 Vsevolod J. Makeev, Dmitry D. Penzar, Ilya E. Vorontsov, Alexander V. Favorov, Julia R. For-  
595 man, Marcia Hasenahuer, Maria S. Fornasari, Gustavo Parisi, Ziga Avsec, Muhammed H. Çelik,  
596 Thi Yen Duong Nguyen, Julien Gagneur, Fang-Yuan Shi, Matthew D. Edwards, Yuchun Guo,  
597 Kevin Tian, Haoyang Zeng, David K. Gifford, Jonathan Göke, Jan Zaucha, Julian Gough, Gra-  
598 ham R. S. Ritchie, Adam Frankish, Jonathan M. Mudge, Jennifer Harrow, Erin L. Young, Yao Yu,  
599 Chad D. Huff, Katsuhiko Murakami, Yoko Nagai, Tadashi Imanishi, Christopher J. Mungall,  
600 Julius O. B. Jacobsen, Dongsup Kim, Chan-Seok Jeong, David T. Jones, Mulin Jun Li, Vi-  
601 oleta Beleva Guthrie, Rohit Bhattacharya, Yun-Ching Chen, Christopher Douville, Jean Fan,  
602 Dewey Kim, David Masica, Noushin Niknafs, Sohini Sengupta, Collin Tokheim, Tychele N.  
603 Turner, Hui Ting Grace Yeo, Rachel Karchin, Sunyoung Shin, Rene Welch, Sunduz Keles, Yue  
604 Li, Manolis Kellis, Carles Corbi-Verge, Alexey V. Strokach, Philip M. Kim, Teri E. Klein, Rahul  
605 Mohan, Nicholas A. Sinnott-Armstrong, Michael Wainberg, Anshul Kundaje, Nina Gonzaludo,  
606 Angel C. Y. Mak, Aparna Chhibber, Hugo Y. K. Lam, Dvir Dahary, Simon Fishilevich, Doron  
607 Lancet, Insuk Lee, Benjamin Bachman, Panagiotis Katsonis, Rhonald C. Lua, Stephen J. Wilson,  
608 Olivier Lichtarge, Rajendra R. Bhat, Lakshman Sundaram, Vivek Viswanath, Riccardo Bellazzi,  
609 Giovanna Nicora, Ettore Rizzo, Ivan Limongelli, Aziz M. Mezlini, Ray Chang, Serra Kim, Car-  
610 men Lai, Robert O’Connor, Scott Topper, Jeroen van den Akker, Alicia Y. Zhou, Anjali D. Zim-  
611 mer, Gilad Mishne, Timothy R. Bergquist, Marcus R. Breese, Rafael F. Guerrero, Yuxiang Jiang,  
612 Nikki Kiga, Biao Li, Matthew Mort, Kymberleigh A. Pagel, Vikas Pejaver, Moses H. Stambou-  
613 lian, Janita Thusberg, Sean D. Mooney, Nuttinee Teerakulkittipong, Chen Cao, Kunal Kundu,  
614 Yizhou Yin, Chen-Hsin Yu, Michael Kleyman, Chiao-Feng Lin, Mary Stackpole, Stephen M.  
615 Mount, Gökçen Eraslan, Nikola S. Mueller, Tatsuhiko Naito, Aliz R. Rao, Johnathan R. Azaria,  
616 Aharon Brodie, Yanay Ofran, Aditi Garg, Debnath Pal, Alex Hawkins-Hooker, Henry Kenlay,  
617 John Reid, Eliseos J. Mucaki, Peter K. Rogan, Jana M. Schwarz, David B. Searls, Gyu Rie Lee,  
618 Chaok Seok, Andreas Krämer, Sohela Shah, ChengLai V. Huang, Jack F. Kirsch, Maxim Shatsky,  
619 Yue Cao, Haoran Chen, Mostafa Karimi, Oluwaseyi Moronfoye, Yuanfei Sun, Yang Shen, Ron  
620 Shigeta, Colby T. Ford, Conor Nodzak, Aneeta Uppal, Xinghua Shi, Thomas Joseph, Sujatha  
621 Kotte, Sadhna Rana, Aditya Rao, V. G. Saipradeep, Naveen Sivadasan, Uma Sunderam, Mario  
622 Stanke, Andrew Su, Ivan Adzhubey, Daniel M. Jordan, Shamil Sunyaev, Frederic Rousseau, Joost  
623 Schymkowitz, Joost Van Durme, Sean V. Tavtigan, Marco Carraro, Manuel Giollo, Silvio C. E.  
624 Tosatto, Orit Adato, Liran Carmel, Noa E. Cohen, Tzila Fenesh, Tamar Holtzer, Tamar Juven-  
625 Gershon, Ron Unger, Abhishek Niroula, Ayodeji Olatubosun, Jouni Väliäho, Yang Yang, Mauno  
626 Vihinen, Mary E. Wahl, Billy Chang, Ka Chun Chong, Inchi Hu, Rui Sun, William Ka Kei Wu,  
627 Xiaoxuan Xia, Benny C. Zee, Maggie H. Wang, Meng Wang, Chunlei Wu, Yutong Lu, Ken  
628 Chen, Yuedong Yang, Christopher M. Yates, Anat Kreimer, Zhongxia Yan, Nir Yosef, Huy-  
629 ing Zhao, Zhipeng Wei, Zhaomin Yao, Fengfeng Zhou, Lukas Folkman, Yaoqi Zhou, Roxana  
630 Daneshjou, Russ B. Altman, Fumitaka Inoue, Nadav Ahituv, Adam P. Arkin, Federica Lovisa,  
631 Paolo Bonvini, Sarah Bowdin, Stefano Gianni, Elide Mantuano, Velia Minicozzi, Leonore No-  
632 vak, Alessandra Pasquo, Annalisa Pastore, Maria Petrosino, Rita Puglisi, Angelo Toto, Liana  
633 Veneziano, Roberta Chiaraluce, Mad P. Ball, Jason R. Bobe, George M. Church, Valerio Con-  
634 salvi, and The Critical Assessment of Genome Interpretation Consortium. Cagi, the critical as-  
635 sessment of genome interpretation, establishes progress and prospects for computational genetic  
636 variant interpretation methods. *Genome Biology*, 25(1):53, February 2024. ISSN 1474-760X.  
637 doi: 10.1186/s13059-023-03113-6.
- 638 Dovini Jayasinghe, Setegn Eshetie, Kerri Beckmann, Beben Benyamin, and S. Hong Lee. Ad-  
639 vancements and limitations in polygenic risk score methods for genomic prediction: a scop-  
640 ing review. *Human Genetics*, 143(12):1401–1431, December 2024. ISSN 1432-1203. doi:  
641 10.1007/s00439-024-02716-8.
- 642 Jin Jin, Bingxuan Li, Xiyao Wang, Xiaochen Yang, Yujue Li, Ruofan Wang, Chenglong Ye, Juan  
643 Shu, Zirui Fan, Fei Xue, Tian Ge, Marylyn D. Ritchie, Bogdan Pasaniuc, Genevieve Wojcik,  
644 and Bingxin Zhao. Pennprs: a centralized cloud computing platform for efficient polygenic  
645 risk score training in precision medicine. *medRxiv: the preprint server for health sciences*, pp.  
646 2025.02.07.25321875, February 2025. doi: 10.1101/2025.02.07.25321875. journalAbbreviation:  
647 medRxiv.
- Chia Hsiang Kao, Evan Trop, McKinley Polen, Yair Schiff, Bernardo P de Almeida, Aaron  
Gokaslan, Thomas PIERROT, and Volodymyr Kuleshov. Advancing dna language models: The

- 648 genomics long-range benchmark. In *ICLR 2024 Workshop on Machine Learning for Genomics*  
649 *Explorations*, 2024.
- 650
- 651 Wolfgang Kopp, Remo Monti, Annalaura Tamburrini, Uwe Ohler, and Altuna Akalin. Deep learning  
652 for genomics using janggu. *Nature Communications*, 11(1):3488, 2020. ISSN 2041-1723. doi:  
653 10.1038/s41467-020-17155-y.
- 654 Kleanthi Lakiotaki, Zaharias Papadovasilakis, Vincenzo Lagani, Stefanos Fafalios, Paulos  
655 Charonyktakis, Michail Tsagris, and Ioannis Tsamardinos. Automated machine learning for  
656 genome wide association studies. *Bioinformatics*, 39(9):btad545, 2023. ISSN 1367-4803. doi:  
657 10.1093/bioinformatics/btad545.
- 658
- 659 Samuel A. Lambert, Laurent Gil, Simon Jupp, Scott C. Ritchie, Yu Xu, Annalisa Buniello, Aoife  
660 McMahon, Gad Abraham, Michael Chapman, Helen Parkinson, John Danesh, Jacqueline A. L.  
661 MacArthur, and Michael Inouye. The polygenic score catalog as an open database for repro-  
662 ducibility and systematic evaluation. *Nature Genetics*, 53(44):420–425, April 2021. ISSN 1546-  
663 1718. doi: 10.1038/s41588-021-00783-5.
- 664 Ko-Han Lee, Yi-Lun Lee, Tsung-Ting Hsieh, Yu-Chuan Chang, Su-Shia Wang, Geng-Zhi Fann,  
665 Wei-Che Lin, Hung-Ching Chang, Ting-Fu Chen, Peng-Husan Li, Ya-Ling Kuo, Pei-Lung Chen,  
666 Hsueh-Fen Juan, Huai-Kuang Tsai, Chien-Yu Chen, and Jia-Hsin Huang. Pgsbuilder: An end-  
667 to-end platform for human genome association analysis and polygenic risk score predictions.  
668 pp. 2023.04.12.536584, April 2023. doi: 10.1101/2023.04.12.536584. URL [https://www.  
669 biorxiv.org/content/10.1101/2023.04.12.536584v1](https://www.biorxiv.org/content/10.1101/2023.04.12.536584v1).
- 670 Zicheng Liu, Jiahui Li, Siyuan Li, Zelin Zang, Cheng Tan, Yufei Huang, Yajing Bai, and Stan Z.  
671 Li. Genbench: A benchmarking suite for systematic evaluation of genomic foundation models.  
672 (arXiv:2406.01627), 2024. doi: 10.48550/arXiv.2406.01627. URL [http://arxiv.org/  
673 abs/2406.01627](http://arxiv.org/abs/2406.01627). arXiv:2406.01627 [q-bio].
- 674
- 675 Timothy Shin Heng Mak, Robert Milan Porsch, Shing Wan Choi, Xueya Zhou, and Pak Chung  
676 Sham. Polygenic scores via penalized regression on summary statistics. *Genetic Epidemiology*,  
677 41(6):469–480, 2017. ISSN 0741-0395. doi: 10.1002/gepi.22050.
- 678 Matthew McDermott. The (lack of?) science of machine learning for healthcare. In *Proceedings*  
679 *of the 4th Machine Learning for Health Symposium*, pp. 19–29. PMLR, February 2025. URL  
680 <https://proceedings.mlr.press/v259/mcdermott25a.html>.
- 681
- 682 Matthew B. A. McDermott, Shirly Wang, Nikki Marinsek, Rajesh Ranganath, Luca Foschini, and  
683 Marzyeh Ghassemi. Reproducibility in machine learning for health research: Still a ways to  
684 go. *Science Translational Medicine*, 13(586):eabb1655, March 2021. ISSN 1946-6242. doi:  
685 10.1126/scitranslmed.abb1655.
- 686 Jorge M. Mendes, Aziz Barbar, and Marwa Refaie. Synthetic data generation: a privacy-preserving  
687 approach to accelerate rare disease research. *Frontiers in Digital Health*, 7, March 2025. ISSN  
688 2673-253X. doi: 10.3389/fdgth.2025.1563991. URL [https://www.frontiersin.org/  
689 journals/digital-health/articles/10.3389/fdgth.2025.1563991/full](https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2025.1563991/full).
- 690 Carene Anne Alene Ndong Sima, Kathryn Step, Yolandi Swart, Haiko Schurz, Caitlin Uren, and  
691 Marlo Möller. Methodologies underpinning polygenic risk scores estimation: a comprehensive  
692 overview. *Human Genetics*, 143(11):1265–1280, November 2024. ISSN 1432-1203. doi: 10.  
693 1007/s00439-024-02710-0.
- 694
- 695 Gherman Novakovsky, Nick Dexter, Maxwell W. Libbrecht, Wyeth W. Wasserman, and Sara  
696 Mostafavi. Obtaining genetics insights from deep learning via explainable artificial intelli-  
697 gence. *Nature Reviews Genetics*, 24(2):125–137, February 2023. ISSN 1471-0064. doi:  
698 10.1038/s41576-022-00532-2.
- 699 Nathan D. Olson, Justin Wagner, Jennifer McDaniel, Sarah H. Stephens, Samuel T. Westreich, An-  
700 ish G. Prasanna, Elaine Johanson, Emily Boja, Ezekiel J. Maier, Omar Serang, David Jáspez,  
701 José M. Lorenzo-Salazar, Adrián Muñoz-Barrera, Luis A. Rubio-Rodríguez, Carlos Flores, Kon-  
stantinos Kyriakidis, Andigoni Malousi, Kishwar Shafin, Trevor Pesout, Miten Jain, Benedict

- 702 Paten, Pi-Chuan Chang, Alexey Kolesnikov, Maria Nattestad, Gunjan Baid, Sidharth Goel,  
703 Howard Yang, Andrew Carroll, Robert Eveleigh, Mathieu Bourgey, Guillaume Bourque, Gen  
704 Li, ChouXian Ma, LinQi Tang, YuanPing Du, ShaoWei Zhang, Jordi Morata, Raúl Tonda, Genís  
705 Parra, Jean-Rémi Trotta, Christian Brueffer, Sinem Demirkaya-Budak, Duygu Kabakci-Zorlu,  
706 Deniz Turgut, Özem Kalay, Gungor Budak, Kübra Narcı, Elif Arslan, Richard Brown, Ivan J.  
707 Johnson, Alexey Dolgoborodov, Vladimir Semenyuk, Amit Jain, H. Serhat Tetikol, Varun Jain,  
708 Mike Ruehle, Bryan Lajoie, Cooper Roddey, Severine Catreux, Rami Mehio, Mian Umair Ahsan,  
709 Qian Liu, Kai Wang, Sayed Mohammad Ebrahim Sahraeian, Li Tai Fang, Marghoob Mohiyud-  
710 din, Calvin Hung, Chirag Jain, Hanying Feng, Zhipan Li, Luoqi Chen, Fritz J. Sedlazeck, and  
711 Justin M. Zook. PrecisionFDA truth challenge v2: Calling variants from short and long reads  
712 in difficult-to-map regions. *Cell Genomics*, 2(5):100129, May 2022. ISSN 2666-979X. doi:  
713 10.1016/j.xgen.2022.100129.
- 714 Oliver Pain, Kylie P. Glanville, Saskia P. Hagenaars, Saskia Selzam, Anna E. Fürtjes, Héléna A.  
715 Gaspar, Jonathan R. I. Coleman, Kaili Rimfeld, Gerome Breen, Robert Plomin, Lasse Folkersen,  
716 and Cathryn M. Lewis. Evaluation of polygenic prediction methodology within a reference-  
717 standardized framework. *PLOS Genetics*, 17(5):e1009021, May 2021. ISSN 1553-7404. doi:  
718 10.1371/journal.pgen.1009021.
- 719 Oliver Pain, Ammar Al-Chalabi, and Cathryn M Lewis. The genopred pipeline: a comprehensive  
720 and scalable pipeline for polygenic scoring. *Bioinformatics*, 40(10):btae551, October 2024. ISSN  
721 1367-4811. doi: 10.1093/bioinformatics/btae551.
- 722 Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor  
723 Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Ed-  
724 ward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner,  
725 Lu Fang, Junjie Bai, and Soumith Chintala. *PyTorch: an imperative style, high-performance deep  
726 learning library*, pp. 8026–8037. Curran Associates Inc., Red Hook, NY, USA, December 2019.  
727
- 728 Aman Patel, Arpita Singhal, Austin Wang, Anusri Pampari, Maya Kasowski, and Anshul Kundaje.  
729 Dart-eval: A comprehensive dna language model evaluation benchmark on regulatory dna. In  
730 *The 38th Conference on Neural Information Processing Systems*, November 2024. URL <https://openreview.net/forum?id=qR0x6H5WUX#discussion>.  
731
- 732 Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier  
733 Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas,  
734 Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duch-  
735 esnay. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.*, 12(null):2825–2830,  
736 November 2011. ISSN 1532-4435.
- 737 Vladislav Perelygin, Alexey Kamelin, Nikita Syzrantsev, Layal Shaheen, Anna Kim, Nikolay Plot-  
738 nikov, Anna Ilinskaya, Valery Ilinsky, Alexander Rakitko, and Maria Poptsova. Deep learning  
739 captures the effect of epistasis in multifactorial diseases. *Frontiers in Medicine*, 11, January 2025.  
740 ISSN 2296-858X. doi: 10.3389/fmed.2024.1479717. URL [https://www.frontiersin.  
741 org/journals/medicine/articles/10.3389/fmed.2024.1479717/full](https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2024.1479717/full).
- 742 Duy Pham, Buu Truong, Khai Tran, Guiyan Ni, Dat Nguyen, Trang T H Tran, Mai H Tran, Duong  
743 Nguyen Thuy, Nam S Vo, and Quan Nguyen. Assessing polygenic risk score models for applica-  
744 tions in populations with under-represented genomics data: an example of vietnam. *Briefings in  
745 Bioinformatics*, 23(6):bbac459, November 2022. ISSN 1477-4054. doi: 10.1093/bib/bbac459.  
746
- 747 Florian Privé, Bjarni J. Vilhjálmsson, Hugues Aschard, and Michael G. B. Blum. Making the most of  
748 clumping and thresholding for polygenic scores. *The American Journal of Human Genetics*, 105  
749 (6):1213–1221, December 2019. ISSN 0002-9297, 1537-6605. doi: 10.1016/j.ajhg.2019.11.001.
- 750 Florian Privé, Julyan Arbel, and Bjarni J. Vilhjálmsson. Ldpred2: better, faster, stronger. *Bioin-  
751 formatics (Oxford, England)*, 36(22–23):5424–5431, April 2021. ISSN 1367-4811. doi:  
752 10.1093/bioinformatics/btaa1029.  
753
- 754 Florian Privé, Julyan Arbel, Hugues Aschard, and Bjarni J. Vilhjálmsson. Identifying and  
755 correcting for misspecifications in gwas summary statistics and polygenic scores. *Human  
Genetics and Genomics Advances*, 3(4), October 2022. ISSN 2666-2477. doi: 10.1016/

- 756 j.xhgg.2022.100136. URL [https://www.cell.com/hgg-advances/abstract/  
757 S2666-2477\(22\)00052-5](https://www.cell.com/hgg-advances/abstract/S2666-2477(22)00052-5).  
758
- 759 Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David  
760 Bender, Julian Maller, Pamela Sklar, Paul I. W. de Bakker, Mark J. Daly, and Pak C. Sham.  
761 Plink: A tool set for whole-genome association and population-based linkage analyses. *American  
762 Journal of Human Genetics*, 81(3):559–575, September 2007. ISSN 0002-9297. doi: 10.1086/  
763 519795.
- 764 Yunfeng Ruan, Yen-Feng Lin, Yen-Chen Anne Feng, Chia-Yen Chen, Max Lam, Zhenglin Guo,  
765 Lin He, Akira Sawa, Alicia R. Martin, Shengying Qin, Hailiang Huang, and Tian Ge. Improving  
766 polygenic prediction in ancestrally diverse populations. *Nature Genetics*, 54(5):573–580, May  
767 2022. ISSN 1546-1718. doi: 10.1038/s41588-022-01054-7.
- 768 Saqib Ul Sabha. *Artificial Intelligence and Omics in Health and Diseases*, pp. 75–92. Springer  
769 Nature, Singapore, 2025. ISBN 978-981-96-8176-1. doi: 10.1007/978-981-96-8176-1\_5. URL  
770 [https://doi.org/10.1007/978-981-96-8176-1\\_5](https://doi.org/10.1007/978-981-96-8176-1_5).  
771
- 772 Castrense Savojardo, Giulia Babbi, Samuele Bovo, Emidio Capriotti, Pier Luigi Martelli, and Rita  
773 Casadio. Are machine learning based methods suited to address complex biological problems?  
774 lessons from cagi-5 challenges. *Human Mutation*, 40(9):1455–1462, 2019. ISSN 1098-1004. doi:  
775 10.1002/humu.23784.
- 776 Max Schuran, Benjamin Goudey, Gillian S Dite, and Enes Makalic. A survey on deep learning for  
777 polygenic risk scores. *Briefings in Bioinformatics*, 26:bbaf373, July 2025. ISSN 1477-4054. doi:  
778 10.1093/bib/bbaf373.  
779
- 780 Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh,  
781 and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based lo-  
782 calization. *International Journal of Computer Vision*, 128(2):336–359, February 2020. ISSN  
783 0920-5691, 1573-1405. doi: 10.1007/s11263-019-01228-7. arXiv:1610.02391 [cs].
- 784 Avanti Shrikumar, Peyton Greenside, and Anshul Kundaje. Learning important features through  
785 propagating activation differences. In *Proceedings of the 34th International Conference on Ma-  
786 chine Learning*, pp. 3145–3153. PMLR, 2017. URL [https://proceedings.mlr.press/  
787 v70/shrikumar17a.html](https://proceedings.mlr.press/v70/shrikumar17a.html).
- 788 Elliot Sollis, Abayomi Mosaku, Ala Abid, Annalisa Buniello, Maria Cerezo, Laurent Gil, Tudor  
789 Groza, Osman Güneş, Peggy Hall, James Hayhurst, Arwa Ibrahim, Yue Ji, Sajo John, Elizabeth  
790 Lewis, Jacqueline A L MacArthur, Aoife McMahon, David Osumi-Sutherland, Kalliope Panou-  
791 sopoulou, Zoë Pendlington, Santhi Ramachandran, Ray Stefancsik, Jonathan Stewart, Patricia  
792 Whetzel, Robert Wilson, Lucia Hindorff, Fiona Cunningham, Samuel A Lambert, Michael In-  
793 ouye, Helen Parkinson, and Laura W Harris. The nhgri-ebi gwas catalog: knowledgebase and  
794 deposition resource. *Nucleic Acids Research*, 51(D1):D977–D985, January 2023. ISSN 0305-  
795 1048. doi: 10.1093/nar/gkac1010.
- 796 Zhan Su, Jonathan Marchini, and Peter Donnelly. Hapgen2: simulation of multiple disease snps.  
797 *Bioinformatics*, 27(16):2304–2305, August 2011. ISSN 1367-4803. doi: 10.1093/bioinformatics/  
798 btr341.  
799
- 800 You Tang and Xiaolei Liu. G2p: a genome-wide-association-study simulation tool for genotype sim-  
801 ulation, phenotype simulation and power evaluation. *Bioinformatics*, 35(19):3852–3854, October  
802 2019. ISSN 1367-4803. doi: 10.1093/bioinformatics/btz126.
- 803 Kimberly A. Tryka, Luning Hao, Anne Sturcke, Yumi Jin, Zhen Y. Wang, Lora Ziyabari, Moira  
804 Lee, Natalia Popova, Nataliya Sharopova, Masato Kimura, and Michael Feolo. Ncbi’s database  
805 of genotypes and phenotypes: dbgap. *Nucleic Acids Research*, 42(D1):D975–D979, January  
806 2014. ISSN 0305-1048. doi: 10.1093/nar/gkt1211.  
807
- 808 Hannah Wand, Samuel A. Lambert, Cecelia Tamburro, Michael A. Iacocca, Jack W. O’Sullivan,  
809 Catherine Sillari, Iftikhar J. Kullo, Robb Rowley, Jacqueline S. Dron, Deanna Brockman, Eric  
Venner, Mark I. McCarthy, Antonis C. Antoniou, Douglas F. Easton, Robert A. Hegele, Amit V.

- 810 Khera, Nilanjan Chatterjee, Charles Kooperberg, Karen Edwards, Katherine Vlessis, Kim Kin-  
811 near, John N. Danesh, Helen Parkinson, Erin M. Ramos, Megan C. Roberts, Kelly E. Ormond,  
812 Muin J. Khoury, A. Cecile J. W. Janssens, Katrina A. B. Goddard, Peter Kraft, Jaqueline A. L.  
813 MacArthur, Michael Inouye, and Genevieve L. Wojcik. Improving reporting standards for poly-  
814 genic scores in risk prediction studies. *Nature*, 591(78497849):211–219, March 2021. ISSN  
815 1476-4687. doi: 10.1038/s41586-021-03243-6.
- 816 Chonghao Wang, Jing Zhang, Werner Pieter Veldsman, Xin Zhou, and Lu Zhang. A comprehen-  
817 sive investigation of statistical and machine learning approaches for predicting complex human  
818 diseases on genomic variants. *Briefings in Bioinformatics*, 24(1):bbac552, January 2023. ISSN  
819 1477-4054. doi: 10.1093/bib/bbac552.
- 820 Sophie Wharrie, Zhiyu Yang, Vishnu Raj, Remo Monti, Rahul Gupta, Ying Wang, Alicia Martin,  
821 Luke J O’Connor, Samuel Kaski, Pekka Marttinen, Pier Francesco Palamara, Christoph Lippert,  
822 and Andrea Ganna. Hapnest: efficient, large-scale generation and evaluation of synthetic datasets  
823 for genotypes and phenotypes. *Bioinformatics*, 39(9):btad535, September 2023. ISSN 1367-4811.  
824 doi: 10.1093/bioinformatics/btad535.
- 825 Menno Witteveen, Emil Pedersen, Joeri Meijssen, Michael Riis Andersen, Florian Privé, Doug  
826 Speed, and Bjarni Vilhjálmsson. Publicly available privacy-preserving benchmarks for polygenic  
827 prediction. October 2022. URL <https://openreview.net/forum?id=ZRZDTY3L8Ne>.
- 828 Ruidong Xiang, Martin Kelemen, Yu Xu, Laura W. Harris, Helen Parkinson, Michael Inouye, and  
829 Samuel A. Lambert. Recent advances in polygenic scores: translation, equitability, methods  
830 and fair tools. *Genome Medicine*, 16(1):33, February 2024. ISSN 1756-994X. doi: 10.1186/  
831 s13073-024-01304-9.
- 832 Sheng Yang and Xiang Zhou. Pgs-server: accuracy, robustness and transferability of polygenic score  
833 methods for biobank scale studies. *Briefings in Bioinformatics*, 23(2):bbac039, March 2022. ISSN  
834 1477-4054. doi: 10.1093/bib/bbac039.
- 835 Tutku Yaraş, Yavuz Oktay, and Gökhan Karakülah. Pgsxplorer: an integrated nextflow pipeline  
836 for comprehensive quality control and polygenic score model development. *PeerJ*, 13:e18973,  
837 February 2025. ISSN 2167-8359. doi: 10.7717/peerj.18973.
- 838 Zijie Zhao, Tim Gruenloh, Meiyi Yan, Yixuan Wu, Zhongxuan Sun, Jiacheng Miao, Yuchang  
839 Wu, Jie Song, and Qiongshi Lu. Optimizing and benchmarking polygenic risk scores with  
840 gwas summary statistics. *Genome Biology*, 25(1):260, October 2024. ISSN 1474-760X. doi:  
841 10.1186/s13059-024-03400-w.
- 842 Zhili Zheng, Shouye Liu, Julia Sidorenko, Ying Wang, Tian Lin, Loic Yengo, Patrick Turley, Alireza  
843 Ani, Rujia Wang, Ilja M. Nolte, Harold Snieder, Jian Yang, Naomi R. Wray, Michael E. Goddard,  
844 Peter M. Visscher, and Jian Zeng. Leveraging functional genomic annotations and genome cov-  
845 erage to improve polygenic prediction of complex traits within and between ancestries. *Nature*  
846 *Genetics*, 56(5):767–777, May 2024. ISSN 1546-1718. doi: 10.1038/s41588-024-01704-y.
- 847 Naihui Zhou, Yuxiang Jiang, Timothy R. Bergquist, Alexandra J. Lee, Balint Z. Kacsoh, Alex W.  
848 Crocker, Kimberley A. Lewis, George Georghiou, Huy N. Nguyen, Md Nafiz Hamid, Larry  
849 Davis, Tunca Dogan, Volkan Atalay, Ahmet S. Rifaioglu, Alperen Dalkıran, Rengul Cetin Atalay,  
850 Chengxin Zhang, Rebecca L. Hurto, Peter L. Freddolino, Yang Zhang, Prajwal Bhat, Fran Supek,  
851 José M. Fernández, Branislava Gemovic, Vladimir R. Perovic, Radoslav S. Davidović, Neven  
852 Sumonja, Nevena Veljkovic, Ehsaneddin Asgari, Mohammad R.K. Mofrad, Giuseppe Profitti,  
853 Castrense Savojardo, Pier Luigi Martelli, Rita Casadio, Florian Boecker, Heiko Schoof, Indika  
854 Kahanda, Natalie Thurlby, Alice C. McHardy, Alexandre Renaux, Rabie Saidi, Julian Gough,  
855 Alex A. Freitas, Magdalena Antczak, Fabio Fabris, Mark N. Wass, Jie Hou, Jianlin Cheng,  
856 Zheng Wang, Alfonso E. Romero, Alberto Paccanaro, Haixuan Yang, Tatyana Goldberg, Chen-  
857 guang Zhao, Liisa Holm, Petri Törönen, Alan J. Medlar, Elaine Zosa, Itamar Borukhov, Ilya  
858 Novikov, Angela Wilkins, Olivier Lichtarge, Po-Han Chi, Wei-Cheng Tseng, Michal Linial, Peter  
859 W. Rose, Christophe Dessimoz, Vedrana Vidulin, Saso Dzeroski, Ian Sillitoe, Sayoni Das,  
860 Jonathan Gill Lees, David T. Jones, Cen Wan, Domenico Cozzetto, Rui Fa, Mateo Torres, Alex  
861 Warwick Vesztrocy, Jose Manuel Rodriguez, Michael L. Tress, Marco Frasca, Marco Notaro, Giu-  
862 liano Grossi, Alessandro Petrini, Matteo Re, Giorgio Valentini, Marco Mesiti, Daniel B. Roche,  
863

864 Jonas Reeb, David W. Ritchie, Sabeur Aridhi, Seyed Ziaeddin Alborzi, Marie-Dominique De-  
865 vignes, Da Chen Emily Koo, Richard Bonneau, Vladimir Gligorijević, Meet Barot, Hai Fang,  
866 Stefano Toppo, Enrico Lavezzo, Marco Falda, Michele Berselli, Silvio C.E. Tosatto, Marco  
867 Carraro, Damiano Piovesan, Hafeez Ur Rehman, Qizhong Mao, Shanshan Zhang, Slobodan  
868 Vucetic, Gage S. Black, Dane Jo, Erica Suh, Jonathan B. Dayton, Dallas J. Larsen, Ashton R.  
869 Omdahl, Liam J. McGuffin, Danielle A. Brackenridge, Patricia C. Babbitt, Jeffrey M. Yunes,  
870 Paolo Fontana, Feng Zhang, Shanfeng Zhu, Ronghui You, Zihan Zhang, Suyang Dai, Shuwei  
871 Yao, Weidong Tian, Renzhi Cao, Caleb Chandler, Miguel Amezola, Devon Johnson, Jia-Ming  
872 Chang, Wen-Hung Liao, Yi-Wei Liu, Stefano Pascarelli, Yotam Frank, Robert Hoehndorf, Maxat  
873 Kulmanov, Imane Boudelloua, Gianfranco Politano, Stefano Di Carlo, Alfredo Benso, Kai  
874 Hakala, Filip Ginter, Farrokh Mehryary, Suwisa Kaewphan, Jari Björne, Hans Moen, Martti E.E.  
875 Tolvanen, Tapio Salakoski, Daisuke Kihara, Aashish Jain, Tomislav Šmuc, Adrian Altenhoff,  
876 Asa Ben-Hur, Burkhard Rost, Steven E. Brenner, Christine A. Orengo, Constance J. Jeffery,  
877 Giovanni Bosco, Deborah A. Hogan, Maria J. Martin, Claire O'Donovan, Sean D. Mooney,  
878 Casey S. Greene, Predrag Radivojac, and Iddo Friedberg. The cafa challenge reports improved  
879 protein function prediction and new functional annotations for hundreds of genes through ex-  
880 perimental screens. *Genome Biology*, 20(1):244, November 2019. ISSN 1474-760X. doi:  
881 10.1186/s13059-019-1835-8.

882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917