

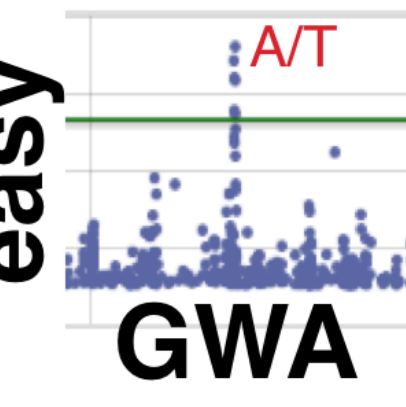
# easyGWAS: A central resource for efficient performance of genome-wide association studies



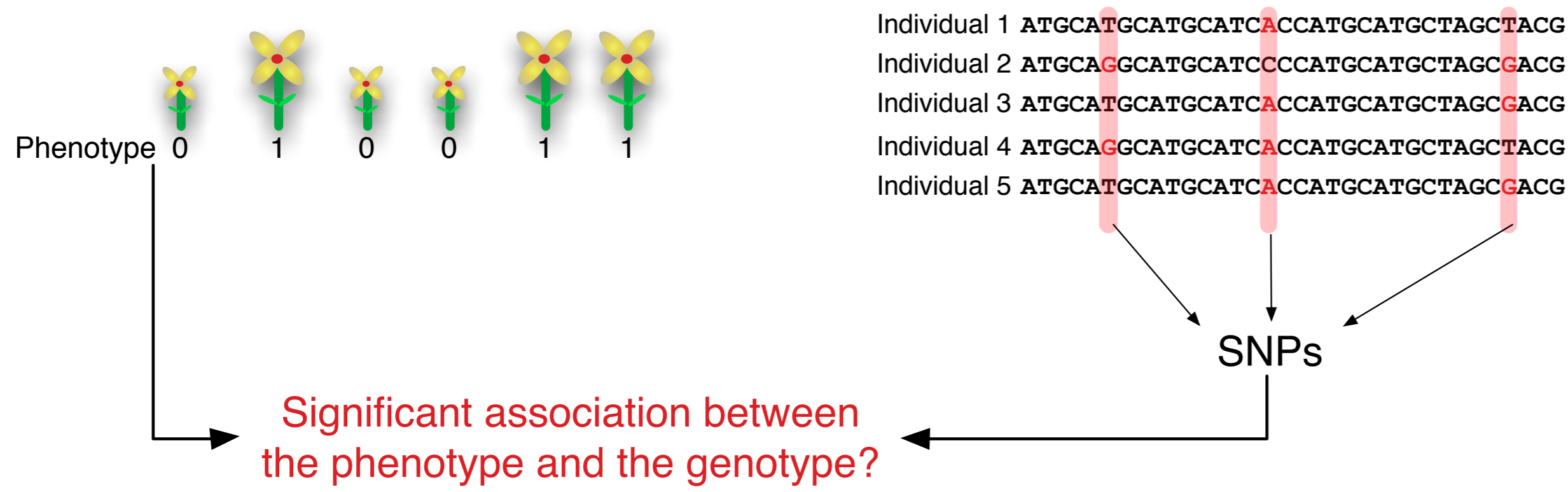
Dominik Grimm<sup>1,2</sup>, Bastian Greshake<sup>1,2</sup>, Oliver Stegle<sup>1,2</sup>, Christoph Lippert<sup>1,2</sup>, Bernhard Schölkopf<sup>2</sup>, Detlef Weigel<sup>1</sup>, Karsten Borgwardt<sup>1,2,3</sup>

Machine Learning and Computational Biology Research Group

<sup>1</sup>Max Planck Institute for Developmental Biology, Tübingen, <sup>2</sup>Max Planck Institute for Intelligent Systems, Tübingen, <sup>3</sup>ZBIT, Eberhard Karls Universität Tübingen



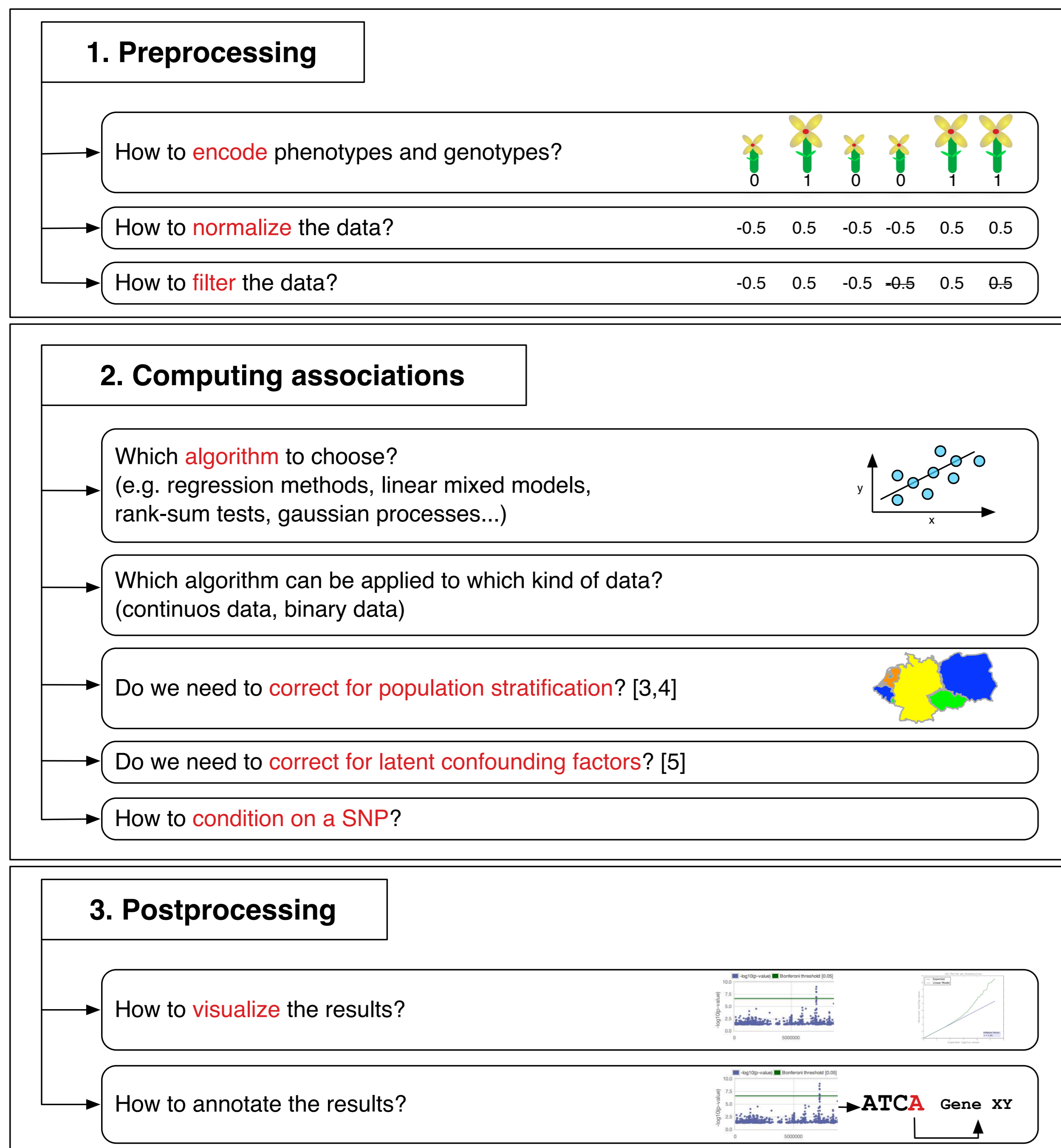
## What is a genome-wide association study?



A genome-wide association (GWA) study examines, whether structural variations such as single nucleotide polymorphisms (SNPs) are associated with a specific phenotypic trait [1, 2]. SNPs are positioned over the whole genome and have a single nucleotide variation. They are tagged within a population of individuals of a certain species. Phenotypes can be observed or measured characteristics. Typically phenotypes are encoded as binary or continuous.

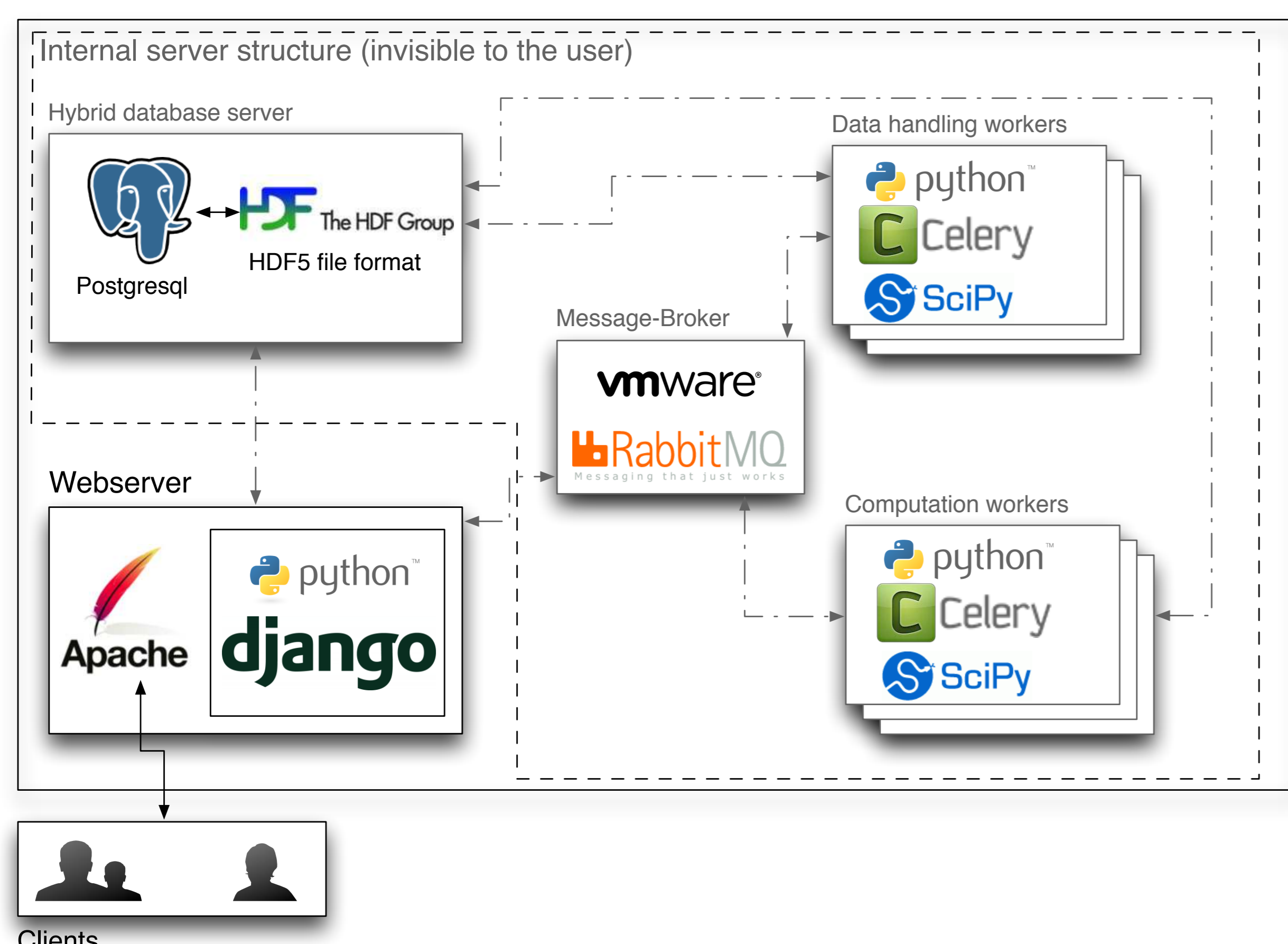
## Performing a GWAS is not trivial

Performing a GWA mapping is **not trivial**, because scientists have to be familiar with many different techniques and methods.



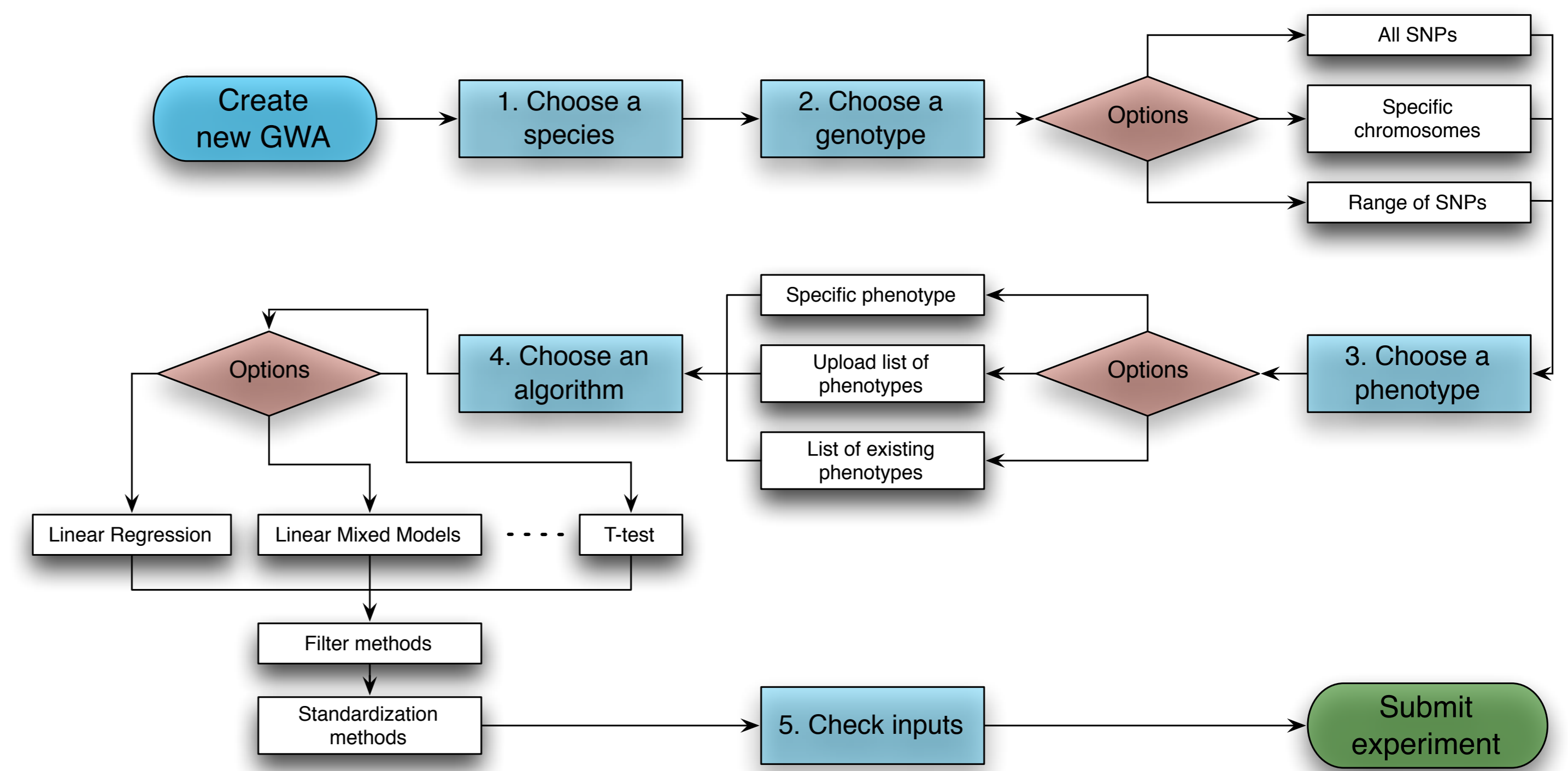
## Performing a GWAS via the web is a challenge

To perform a GWA mapping a huge amount of data has to be analyzed and handled. For that purpose, many computational resources and intelligent data handling methods are essential. Furthermore, to manage and compute GWA studies for different users, smart scheduling techniques have to be used to ensure an unobstructed workflow.



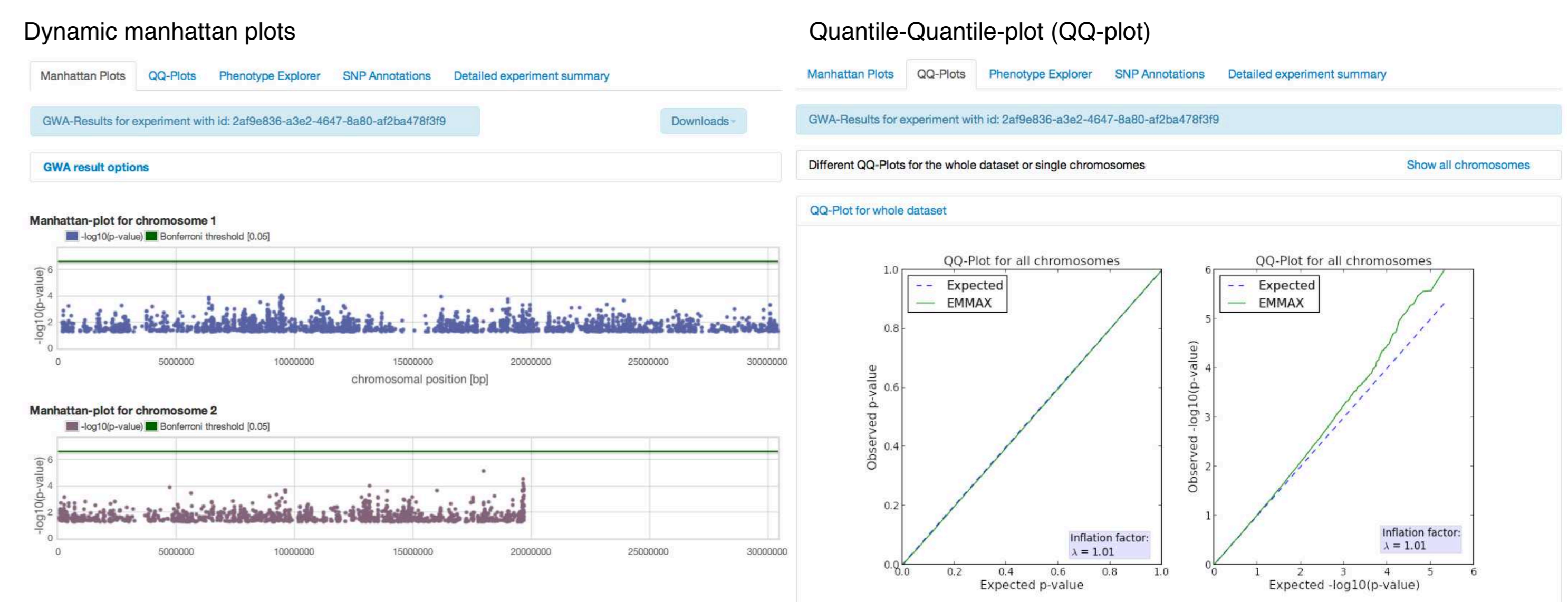
## The easyGWA experiment wizard

The **easyGWA** experiment wizard is a step-by-step procedure to easily create new GWA study via the web.



## The easyGWA result view

To present the results of GWA experiments we provide different visualizations. Manhattan-plots show at the y-axis minus log transformed p-values and on the x-axis the genomic position. The quantile-quantile-plots (QQ-plots) compare the expected null-distribution of a sorted set of single p-values with the observed distribution of a sorted set of all computed p-values.



## Providing and sharing results with other scientists

All the results of a computed GWA study can be downloaded in different formats. Every user has the ability to run GWA experiments in a **private environment**. The users have the possibility to make significant result public to the scientific community. Furthermore, users can **share** and forward their results to other members via the web-interface.

## The easyGWA data browser

To gather experiences and detailed information about the integrated data the user can browse the data. All genotypes and phenotypes are tagged with meta-informations such as the longitude and latitude which are displayed using Google-Maps<sup>®</sup>. To retrieve specific data the user can build his own queries with an easy-to-use query builder.

## Summary

- **Easily** create new GWA-experiments via the web using different methods and covariates
- **Visualize, analyze** and **annotate** GWA results
- **Private** and **public** instances of phenotypic data
- **Share** your GWA results with your colleges or make them public
- **Download** and browse genotypic and phenotypic data
- **Future directions:**
  - Upload **own genotypic** data for private analysis
  - **Multi-marker** and **multi-locus** search

## Acknowledgements

- Barbara Rakitsch and Stefan Kleeberger

## References

- [1] Susanna Atwell, Yu S. Huang, and Bjarni J. et al. Vilhjalmsjon. Genome-wide association study of 107 phenotypes in arabidopsis thaliana inbred lines. *Nature*, 465(7298):627–631, June 2010.
- [2] Joel N. Hirschhorn and Mark J. Daly. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*, 6(2):95–108, February 2005.
- [3] Christoph Lippert, Jennifer Listgarten, Ying Liu, Carl M Kadie, Robert I Davidson, and David Heckerman. FaST linear mixed models for genome-wide association studies. *Nat Meth*, 8(10):833–835, October 2011.
- [4] Alkes L Price, Nick J Patterson, Robert M Plenge, Michael E Weinblatt, Nancy A Shadick, and David Reich. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*, 38(8):904–909, 2006.
- [5] Nicol Fusi, Oliver Stegle, and Neil D. Lawrence. Joint modelling of confounding factors and prominent genetic regulators provides increased accuracy in genetical genomics studies. *PLoS Comput Biol*, 8(1):e1002330, January 2012.