



# The Mega2R R package: Tools for accessing and processing common genetic data formats in R.

Daniel E. Weeks<sup>1,2</sup>, Robert V. Baron<sup>1</sup>, Justin R. Stickel<sup>1</sup>

<sup>1</sup>Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, <sup>2</sup>Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

## Mega2

### What is Mega2:

Data reformatting tool in C++ for genetic analysis<sup>1,2</sup>

### Input formats:

|                 |         |
|-----------------|---------|
| PLINK PED/BED   | VCF/BCF |
| Linkage         | Impute2 |
| Mega2 Annotated |         |

### Output formats/analysis programs: (Recently added)

|                  |          |
|------------------|----------|
| Minimac3/Shapeit | VCF      |
| ROADTRIPS        | PLINK    |
| Loki             | Mendel   |
| PLINK/SEQ        | FBAT     |
| Cranefoot        | Morgan   |
| Merlin           | Beagle   |
| Mega2 Annotated  | Simwalk2 |

### Recent enhancements:

- SQLite3 Database
- Alignment to Reference Alleles

### SQLite3 Database Tables:

- Pedigree/Person tables
- Locus/Marker/Allele/Map tables
- Phenotype/Genotype tables

### Availability:

- Open source (C++) implementation.
- Freely available.
- Linux, Mac, Solaris, and Windows

<https://watson.hgen.pitt.edu/register/>



Mega2->



Mega2R->

### References:

1. Baron et al. (2014) Mega2: validated data-reformatting for linkage and association analyses. *Source Code Biol Med.* 2014 Dec 5;9(1):26.
2. Mukhopadhyay et al. (2005) Mega2: data-handling for facilitating genetic linkage and association analyses. *Bioinformatics* 21:2556-255.
3. Schaid et al. (2013) Multiple genetic variant association testing by collapsing and kernel methods with pedigree or population structured data. *Genet Epidemiology* 37(5):408-418.
4. Aulchenko YS et al. (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics*. 2007 May 15;23(10):1294-6.
5. Wu MC et al. (2011) Rare Variant Association Testing for Sequencing Data Using the Sequence Kernel Association Test (SKAT). *Am J Hum Genet*, 89, 82-93.

**Acknowledgments:** This work was supported by NIH grant R01 GM076667 (PI: Weeks)

## Mega2R

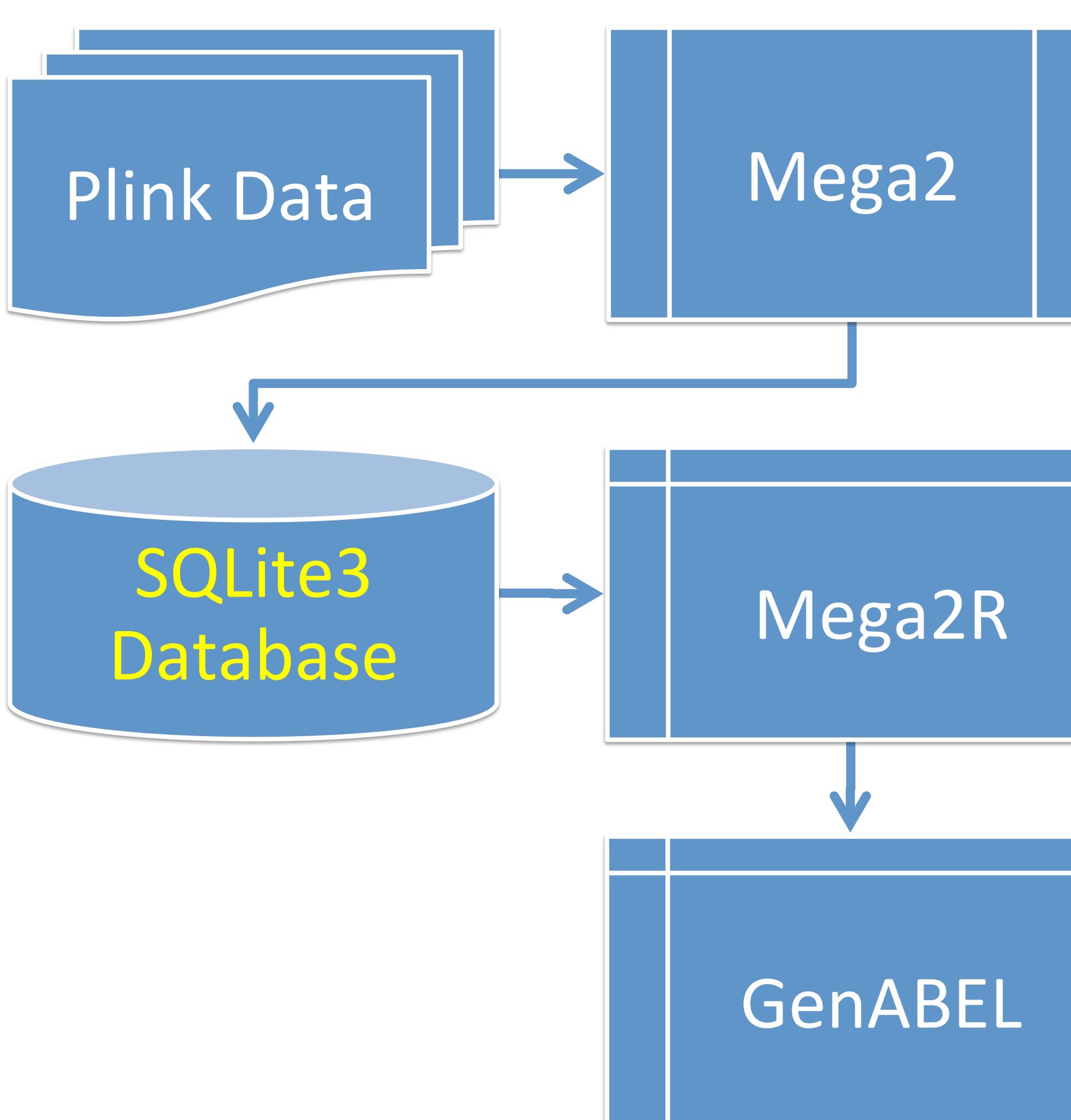
### What is Mega2R:

An efficient pathway from many data formats via a **Mega2-created database** into R data frames. These data can be easily used in other R packages and functions.

### Input:

Mega2 SQLite3 database tables.

### Example conversion to GenABEL:



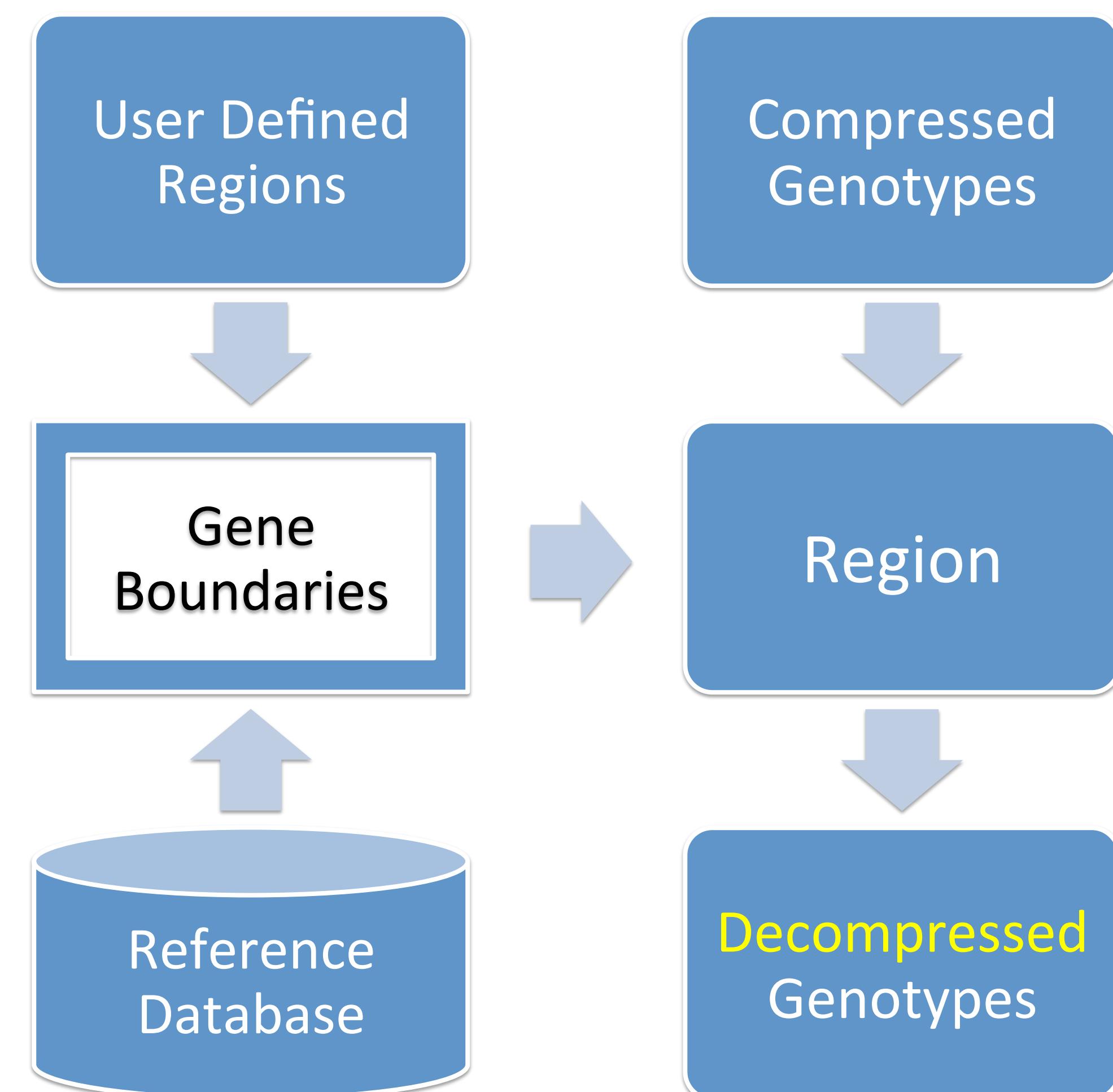
### Region extraction:

**Memory-efficient extraction** of ranges of genotype data within selected regions.

Regions can be defined by the user or by using a transcription database such as those available in Bioconductor.

**Genotype data is compressed** in the database for storage purposes, Mega2R can extract and decompress data for use in Mega2R's iterative wrapper or for any other R packages.

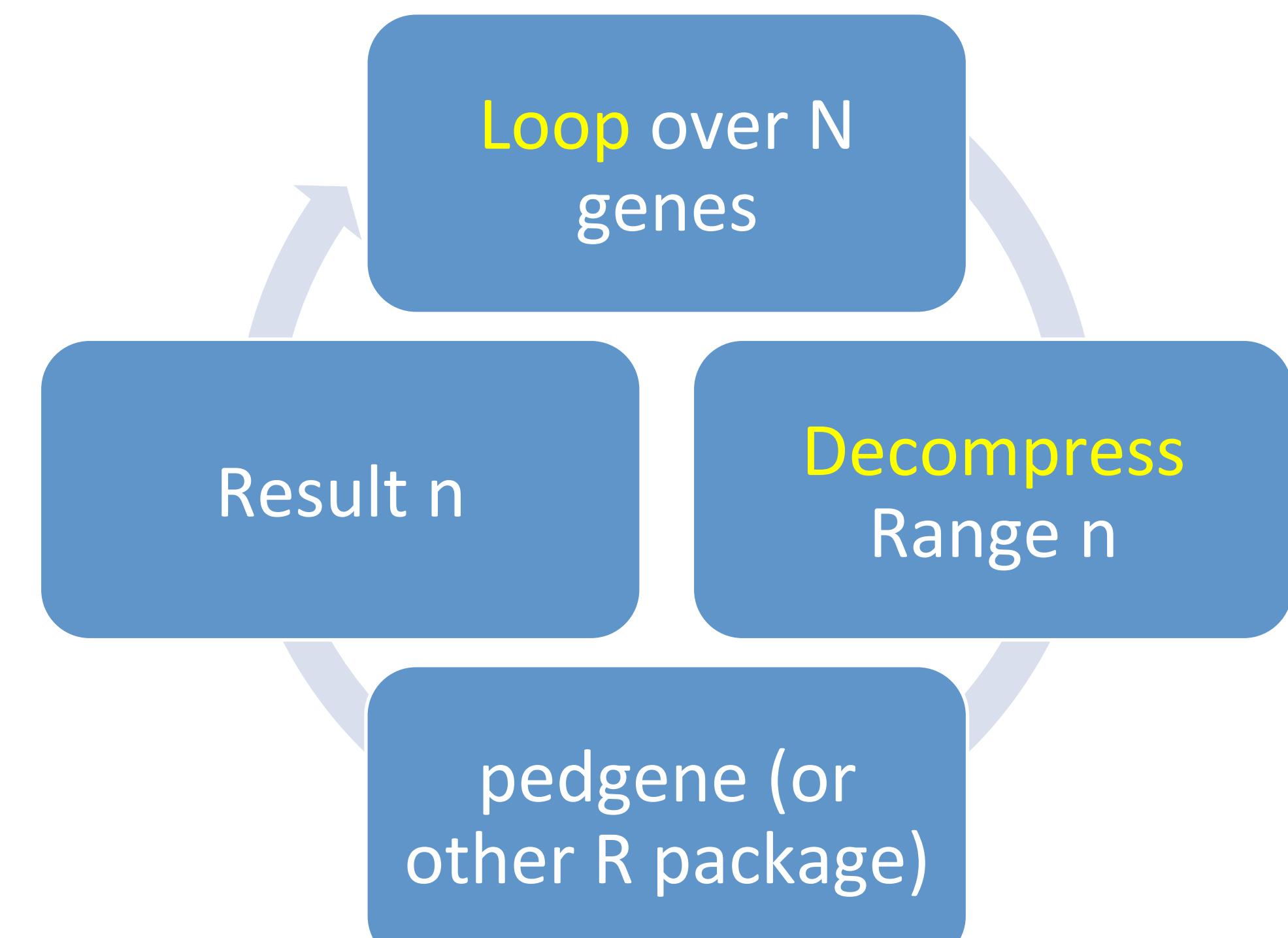
### Region extraction diagram:



### Iteration wrapper:

**Extensive efficient flexible wrapper** to iterate through gene regions.

This wrapper can run R functions, which analyze only a single region at a time, on multiple genes.



### Example R functions:

- Mega2pedgene<sup>3</sup>
- Mega2GenABEL<sup>4</sup>
- Mega2VCF
- Mega2SKAT<sup>5</sup>

**Requirements:** R, Mega2

**Available with tutorial:** CRAN

<https://watson.hgen.pitt.edu/mega2/mega2r/>