

# pheno2geno

Konrad Zych<sup>1,2</sup>, Danny Arends<sup>2</sup>, Ritsert C. Jansen<sup>2</sup>

June 20, 2011

1. Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University in Krakow, Poland
2. Groningen Bioinformatics Center, University of Groningen, The Netherlands

# 1 General introduction

Genetical Genomics (citation) is powerful method, providing world of life sciences with tool to look deep inside the complex relation between genetic information stored by DNA and final outcome of its processing - phenotype. And because this is the core dogma of modern biology, every method helping us to understand it better is of high importance to scientific world.

Great power often comes for high price, though. To perform GG studies one has to obtain both phenotypic and genotypic data, that are subsequently being matched. This not only elevates costs of experiment, but also introduces number of human errors, e.g. mismatching/mislabeling of arrays.

This brought us back to DNA to phenotype dogma. And we came up with idea of creating genetic map out of gene expression data. This means the same power for less than half of the cost and effort. Procedure is easy enough to be conducted by inexperienced R user and for advanced users we offer variety of extra functionalities to make their analysis fit their needs.

## 1.1 R programming language

R programming language is powerful, yet easy-to-use. There is graphical interface available for Windows, Mac OS and Linux. Every package/function comes with easily accessible help file and, most importantly, R provides user with handfulls of statistical functionalities. To start your adventure with R just go to: <http://cran.r-project.org/>, select your operating system, install it, and you're ready to enter the world of R.

## 1.2 Downloading and installing package

R packages contribute to power of this language more than anything. Using then, you can extend basic R to powerful tool perfectly suited to your personal needs. Installing package is really easy, just open R gui and type: `install.me` or what? or use Packages menu (just click on Install Packages, select a mirror and than select package pheno2geno).

## 2 Data

### 2.1 Data files' structure

In order to ensure smooth work of our package, we specified strict rules about data files provided. If below mentioned requirements are met and filenames are default, user don't have to provide reading function with any parameters. If your more advanced R user, your data files are of different format or you have data already inside R environment, please see section 4.1.

#### 2.1.1 Phenotype data for offspring

This data is crucial and analysis cannot be run without it. Default filename is `offspring_phenotypes.txt`. Rows correspond to markers and columns to individuals. File should contain only numeric values apart from first row, containing unique column names and first column containing unique row names. Rows and columns with non-numeric values are removed. All values should be separated by tabs. In normal practice first row has one element less then others (no column name for column containing rownames).

#### 2.1.2 Phenotype data for founders

This data is optional but really important for the analysis, so we recommend to provide it. Default filename is `founders_phenotypes.txt`. Data structure inside the file should be the same as in offspring phenotype file. Also

### 2.2 Population object

To store all the necessary information and mainupulate it easily we developed object of class `population` (schematically shown on figure 1). It's divided in three main parts:

- founders part, storing founders phenotype data, information about founders groups used by RankProduct analysis (see 3.2) and results of this analysis (RP object, futher documentation about it is available in help file of RP function),
- offspring part, storing founders phenotype and genotype data (read from file and/or simulated by our software),
- maps part, storing genetic and/or physical map

row names should match, because non-matching rows are removed.

#### 2.1.3 Genotype data for offspring

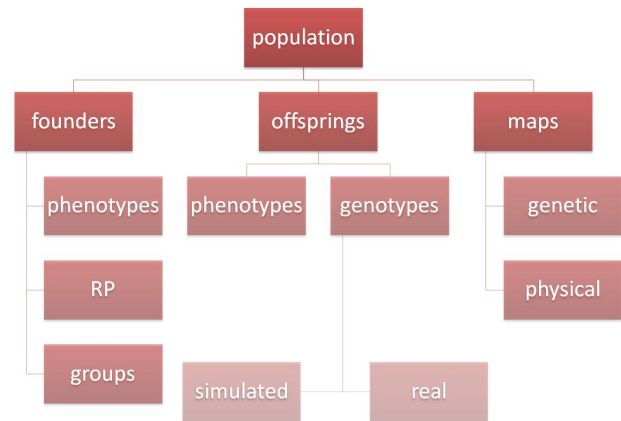
This data is optional and doesn't have any impact on analysis, could be only used for creating cross object out of existing data using our software. Default filename is `offspring_genotypes.txt`. Genotypes should be coded as 0,1 and NA. Data structure inside the file should be the same as in offspring and founders phenotype files. Row names should match, because non-matching rows are removed.

#### 2.1.4 Genetic map

This data is optional and could be used for comparison with map created by our software or for ordering markers in output cross. Default filename is `maps_genetic.txt`. File should contain three columns - marker name, chromosome number and position on chromosome in cM. Second and third column should contain numeric values without NAs, NaNs, etc. First column should contain unique marker names, matching ones in phenotype files.

#### 2.1.5 Physical map

Exactly the same as in genetic map, but position on chromosome should be in Mbp.



## 3 Basic workflow

### 3.1 Loading data into workflow

If data is formatted as specified in chapter 2 and files are having default names, you just have to type: `population` `j- readFiles()` into Rgui (make sure your working directory is the one containing data files) and it's all done. If you want to customize your files names, that can be done, but only to some extent. First of all you're able only to specify "cores":

- founders core, then name of the file containing founders phenotype data should be `founders core_phenotypes.txt`
- offspring core, then name of the file containing offspring phenotype data should be `offspring core_phenotypes.txt` and file containing genotype data should be called `offspring core_genotypes.tx`
- maps core, then file `maps core_genetic.txt` contains genetic map, while `maps core_physical.txt` - physical map.

If you are more advanced in R and want to use your own data files of different names and/or structure, see section 4.1.

### 3.2 Rank product analysis

Rank Product (citation) is a powerful method for discovery of

### 3.3 Preoptimized parameters for most common experimental crosses

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

### 3.4 Selecting appropriate markers

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters

of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

### 3.5 Splitting selected markers

#### 3.5.1 Parental mean splitting

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

#### 3.5.2 EM splitting

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

### 3.6 Filtering markers

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

### 3.7 Cross object

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift –

Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

### **3.7.1 Creating cross object**

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

### **3.7.2 Forming linkage groups and ordering markers**

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place.

If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

### **3.7.3 Augmenting cross object**

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

## 4 Advanced options/modifications

### 4.1 Using data files with different structure

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

### 4.2 Rank product analysis

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

### 4.3 Uncommon types of crosses

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

### 4.4 Modifying splitting options

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

### 4.5 Filtering markers

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

### 4.6 Cross object

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

#### 4.6.1 Forming linkage groups and ordering markers

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

#### 4.6.2 Post-processing of cross object

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

## 5 Built-in plotting routines

### 5.1 plotChildrenExpression

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language. Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

### 5.2 plotParentalExpression

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language. Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

### 5.3 plotMapComparison

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language. Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

### 5.4 plotMarkerDistribution

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language. Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

## 6 Big datasets

### 6.1 Problematic handling of big data by R

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

### 6.2 C preprocessing

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift –

Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.

### 6.3 Other solutions

Hello, here is some text without a meaning. This text should show, how a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like »Huardest gefburn«. Kjift – Never mind! A blind text like this gives you information about the selected font, how the letters are written and the impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for a special contents, but the length of words should match to the language.



## 7 Package development & collaboration

## 8 References