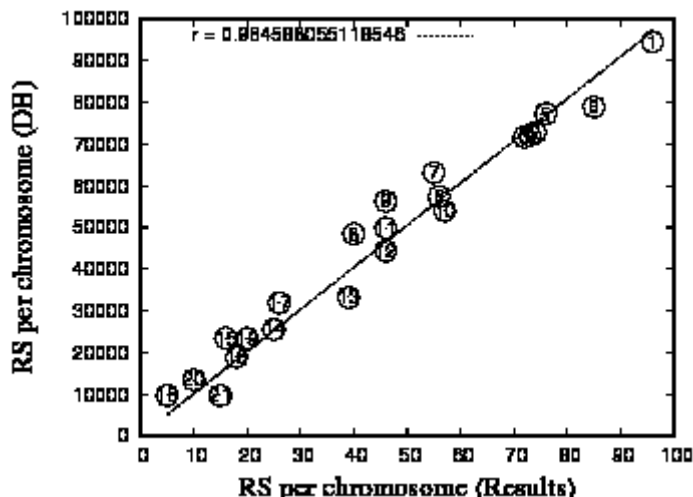


Comparing results from different methods

The goal is to provide a method to identify how good is a result without further information. The basic idea behind it is that a set of random SNPs from a database must fulfill a very good linear regression against this database.

As an example, in the next picture a linear regression of this kind is shown.

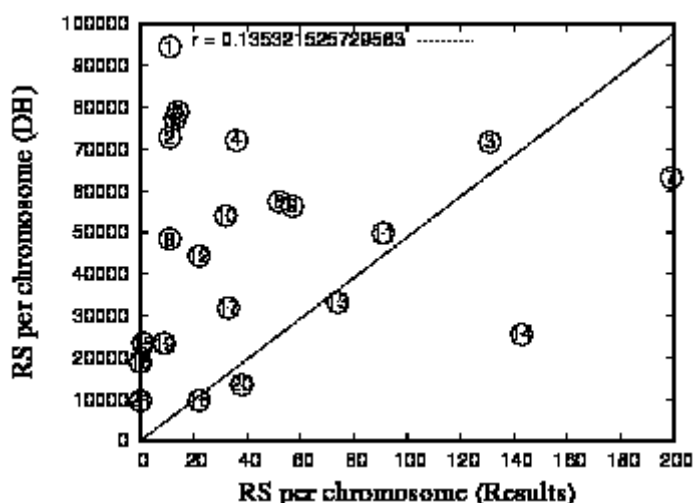


Here a regression is calculated with, as usual,

$$\beta = \frac{\overline{xy} - \overline{x}\overline{y}}{\overline{x^2} - \overline{x}^2}$$

$$r = \frac{n \sum \{x_i y_i\} - \sum \{x_i\} \sum \{y_i\}}{\sqrt{\left(n \sum \{x_i^2\} - \left(\sum \{x_i\} \right)^2 \right) \left(n \sum \{y_i^2\} - \left(\sum \{y_i\} \right)^2 \right)}}$$

However, if a set of SNPs are not random distributed the regression must performs bad, as the one shown in next figure.



But there is a problem with this method. It is straightforward that if the regression is good the SNPs are random distributed and results are not representative of nothing more than data randomness. But this is not necessarily true in the opposite way. That is, a low pearson correlation does not imply the

SNPs are right chosen. **This just implies they are not random!**

Practical Guides

See [here](#) and [here](#)

From:

<https://imagen.fundacioace.com/wiki/> - **Detritus Wiki**

Permanent link:

https://imagen.fundacioace.com/wiki/doku.php?id=genetica:comp_histo

Last update: **2020/08/04 10:58**

