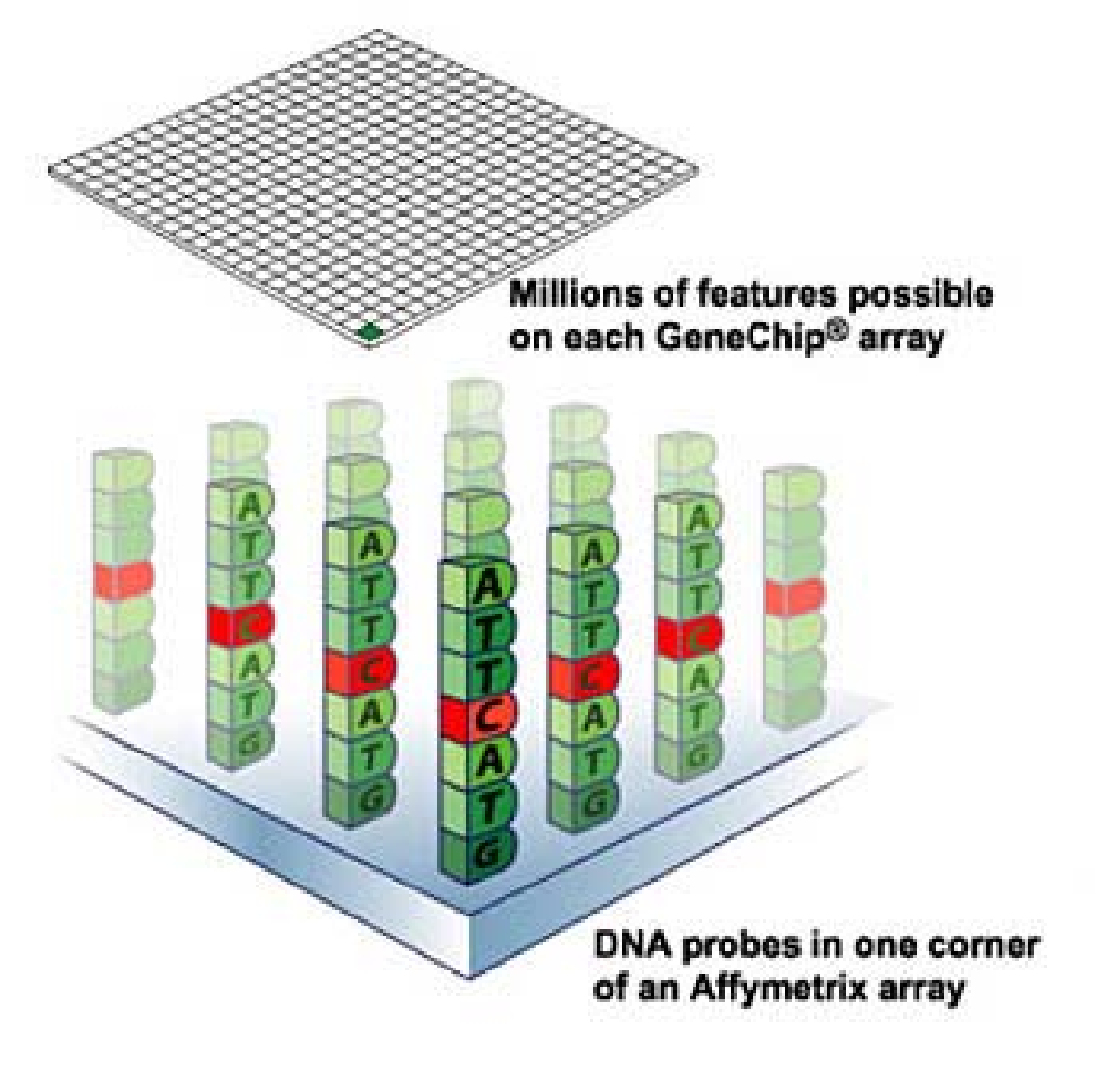
Multiple testing etc.

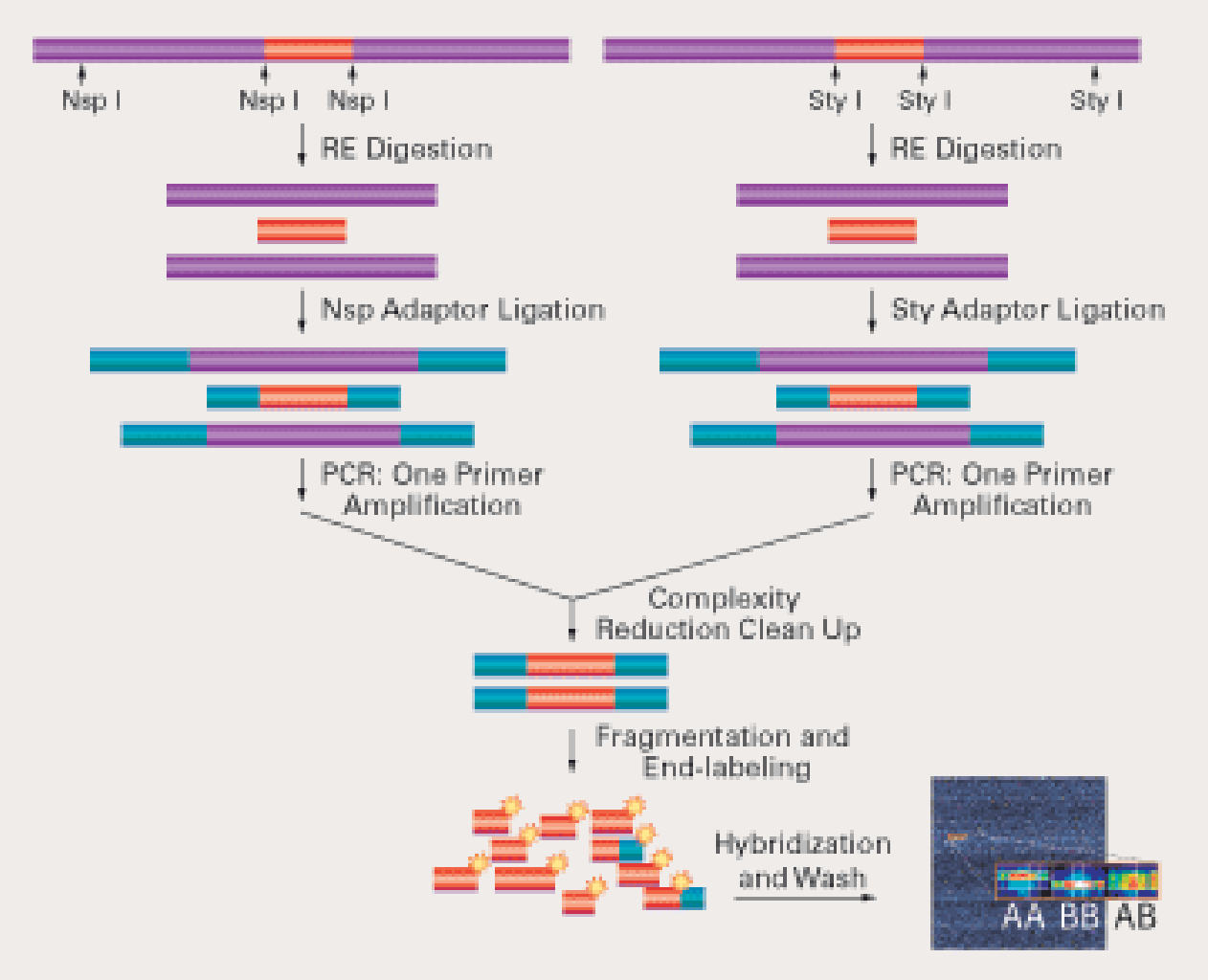
Benjamin Neale

Leuven 2008

Technology advances – Affy 500K



How the Affy Chip works



Illumina's Chip

Multiple testing approaches

- Bonferroni Correction
  - Conservative
  - Probability of observing at least 1 hit at that level
- False Discovery methodology
  - Capitalizes on a distribution of positive results
  - See work by Benjamin, Hochberg, Storey
- Bayes Factor
  - Like a P-value, but for Bayesians
  - Marginal likelihood of the two sets of parameters
  - Correlation approaches 1 under the alternative

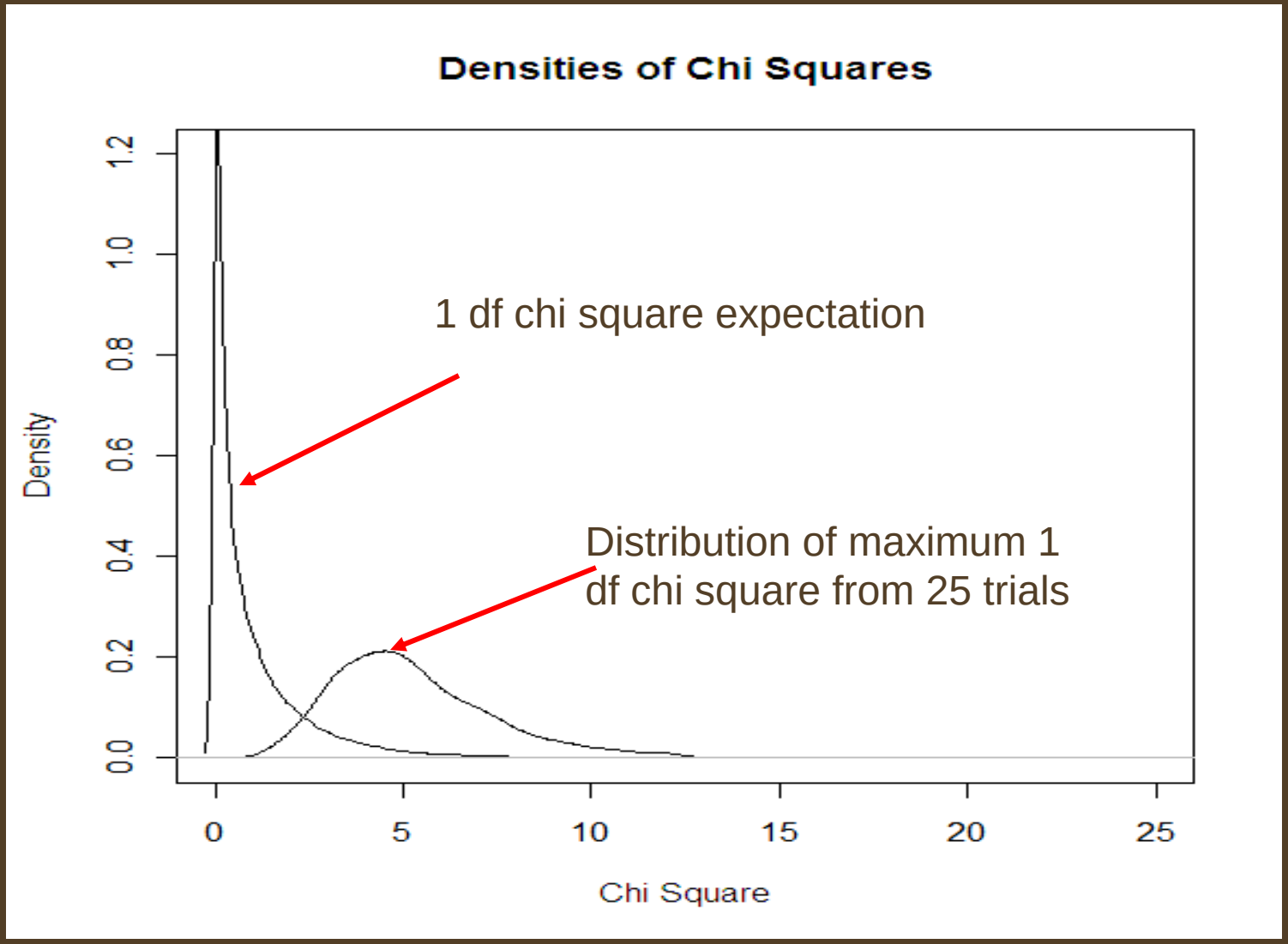
Genome-wide significance

- Multiple testing theory requires an estimate of the number of ‘independent tests’
- Risch and Merikangas 1996 estimated a threshold of 10^{-6} = (0.05/(5*10000))
- HapMap 2005 estimate 10^{-8} based on encode deep sequencing in ENCODE regions
- Dudbridge and Gusnato and Pe’er et al. 2008 Genetic Epidemiology estimate based on ‘infinite density’ like Lander and Kruglyak 1995 generates $5x10^{-8}$

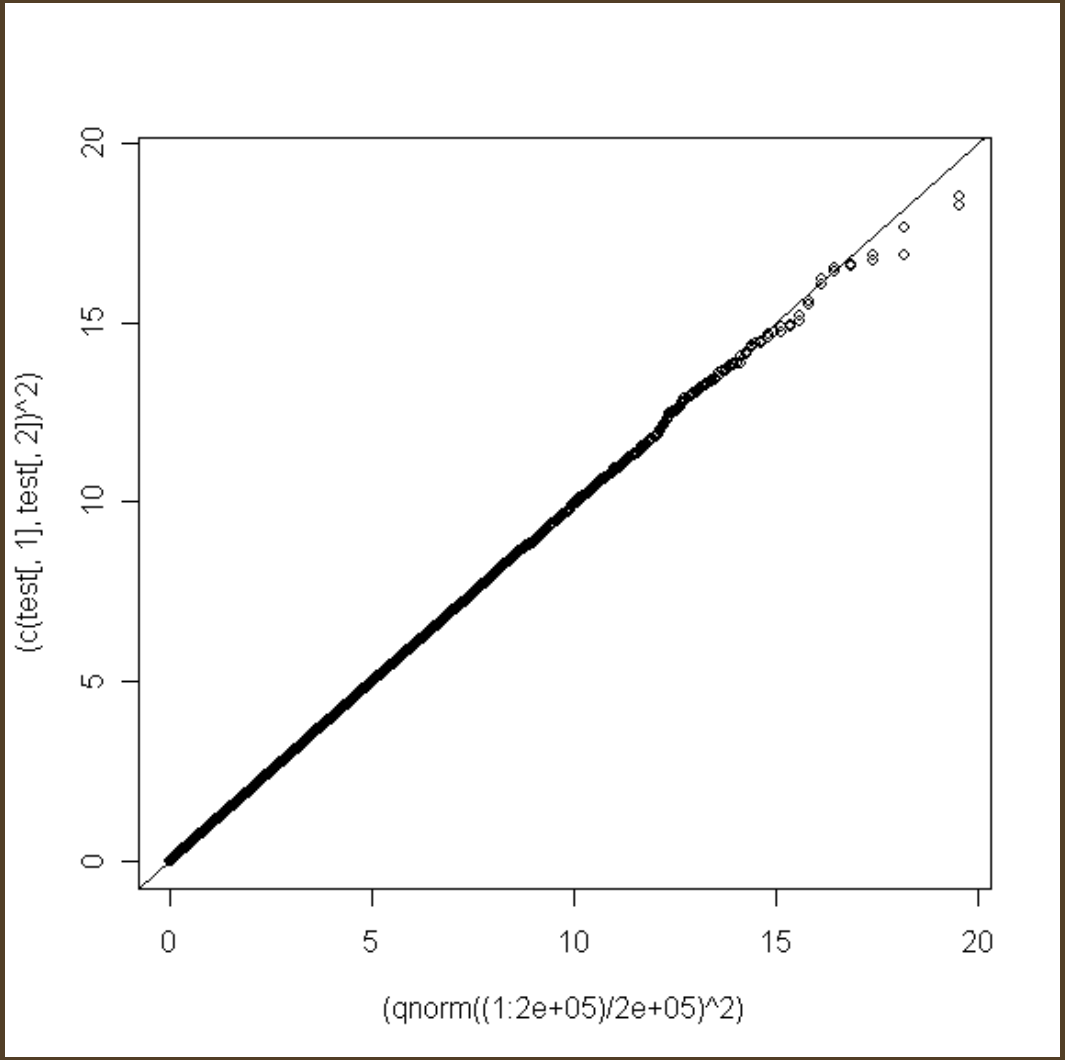
Guessing game

- I’ll buy a beer for the closest guess
- We’re each going to simulate a single chi square
- There are about 25 groups

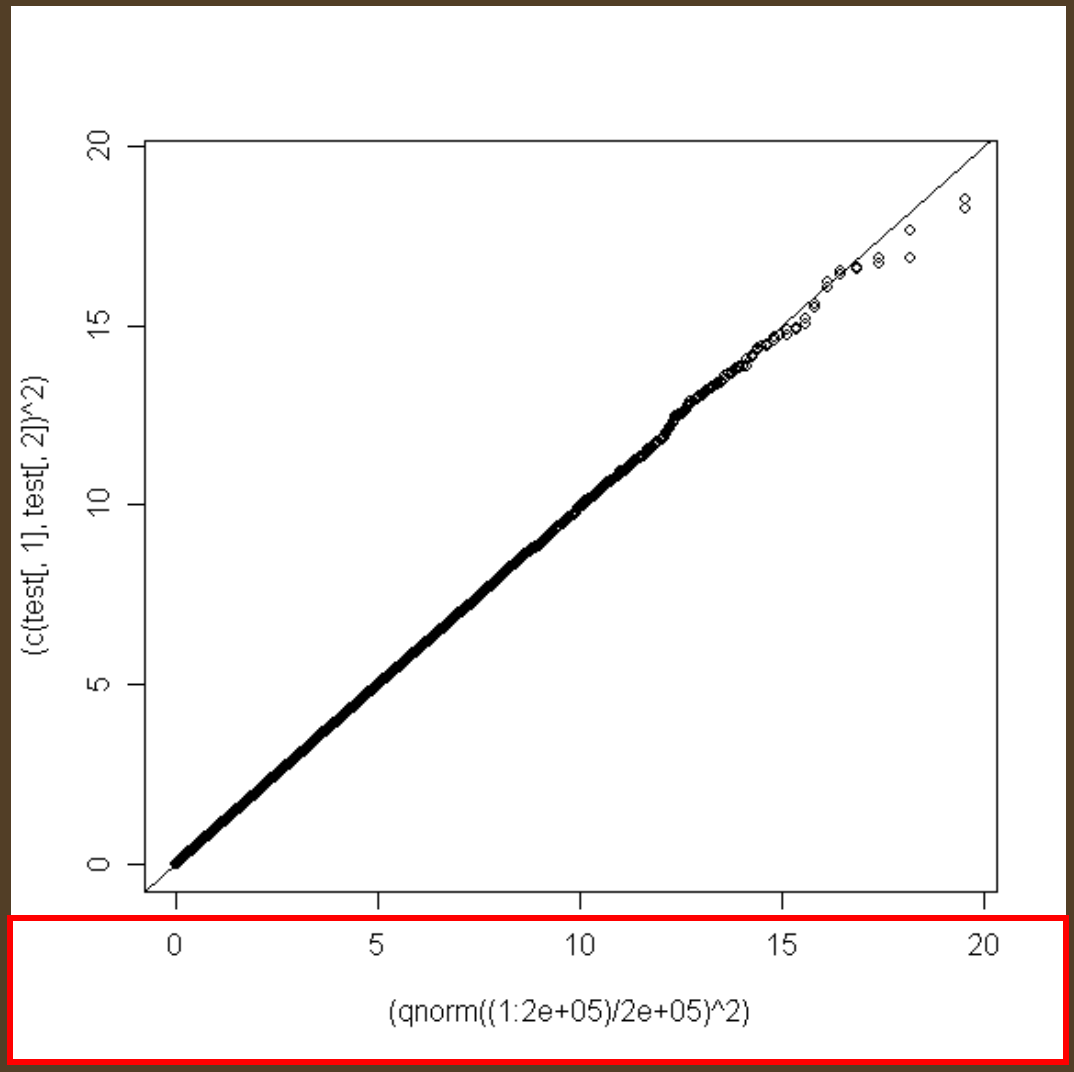
Distribution of chi squares



QQ plotting



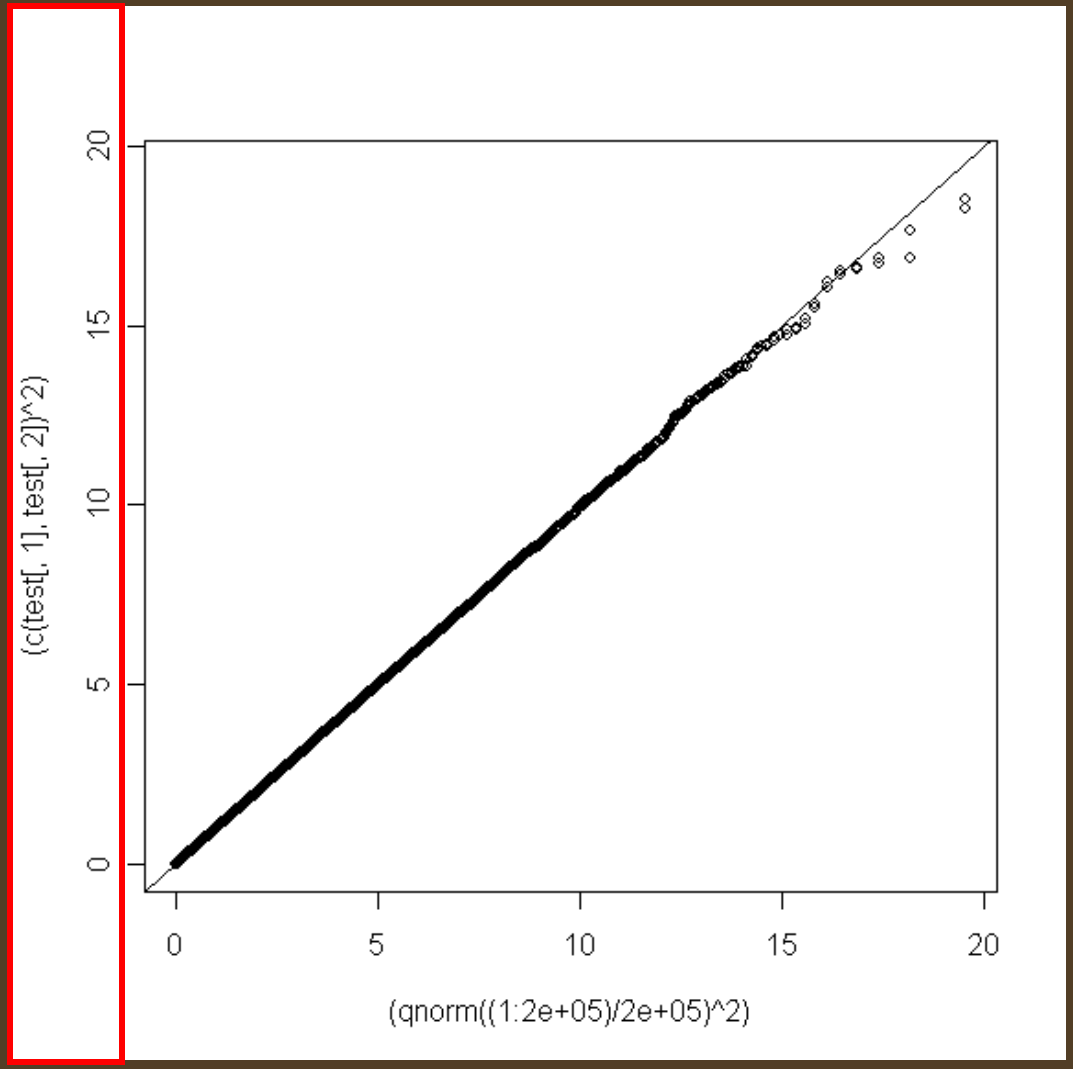
QQ plotting



Expectation

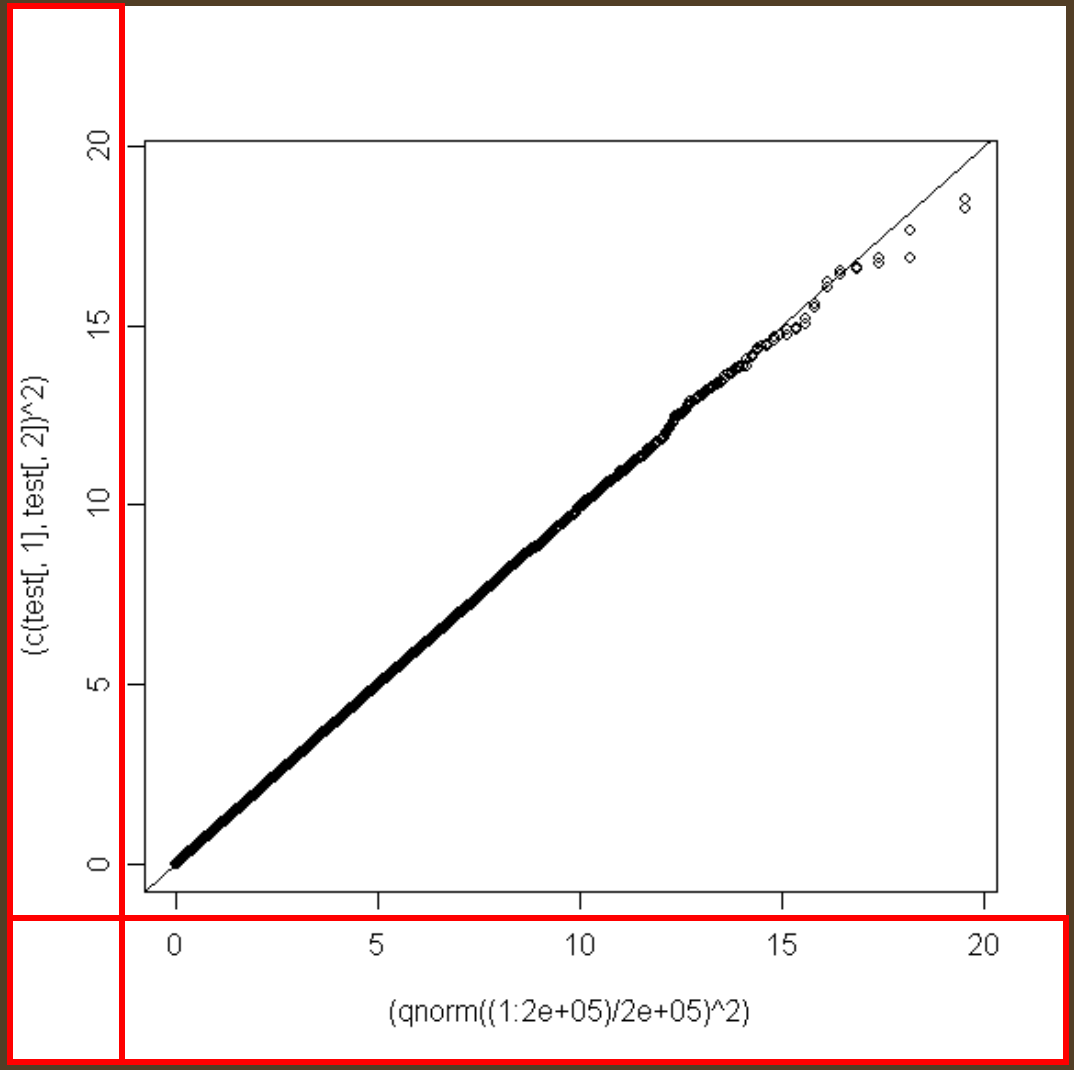
QQ plotting

Observation



QQ plotting

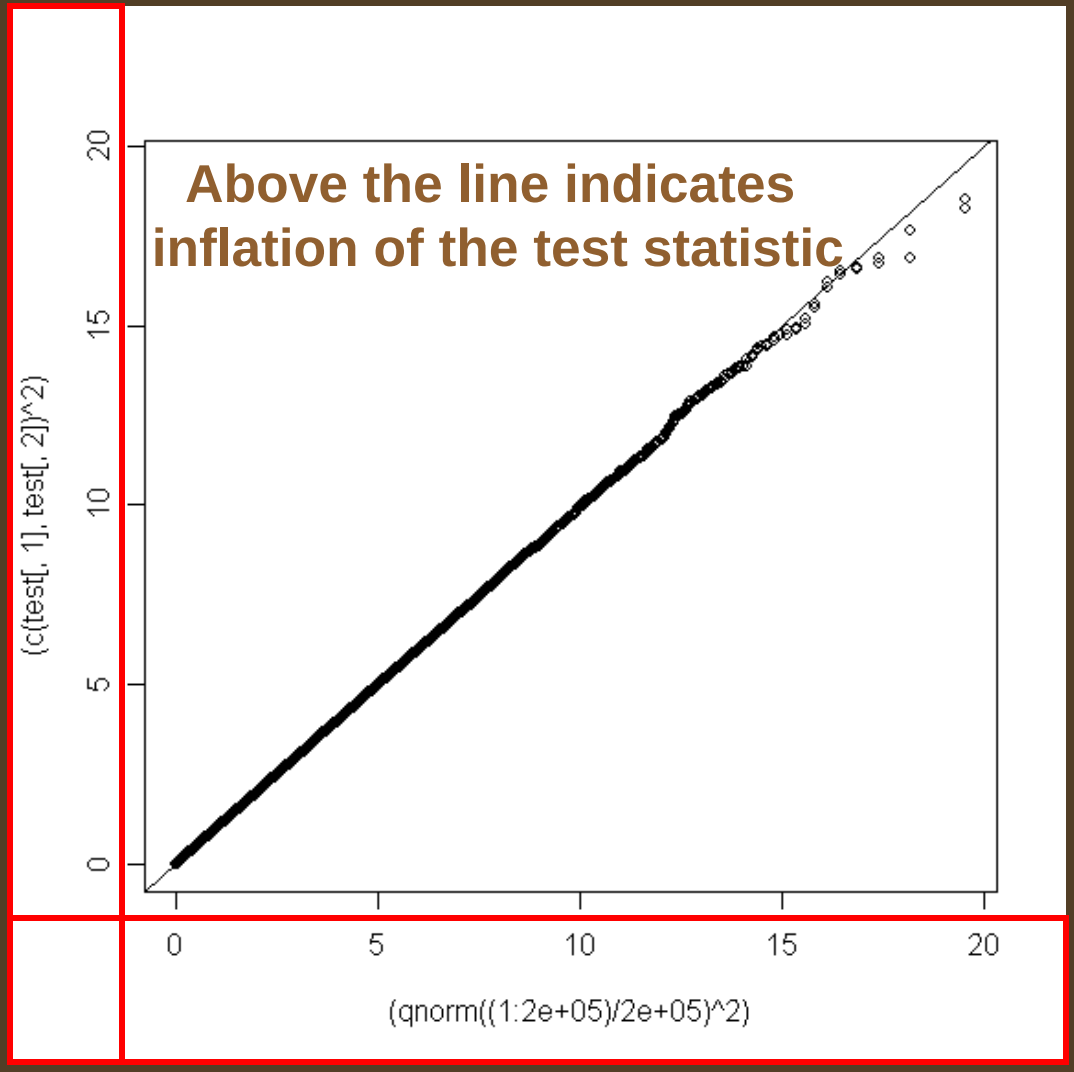
Observation



Expectation

QQ plotting

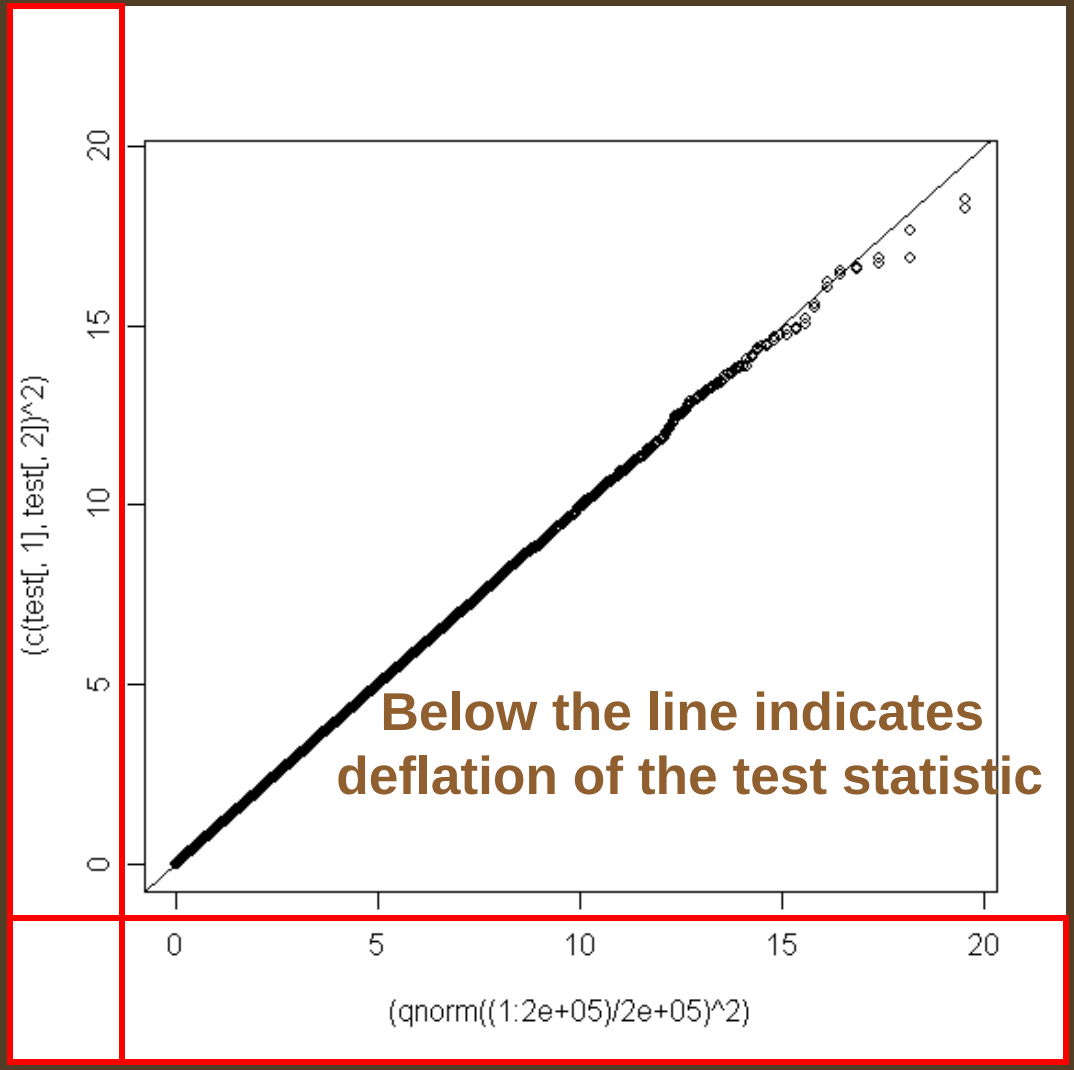
Observation



Expectation

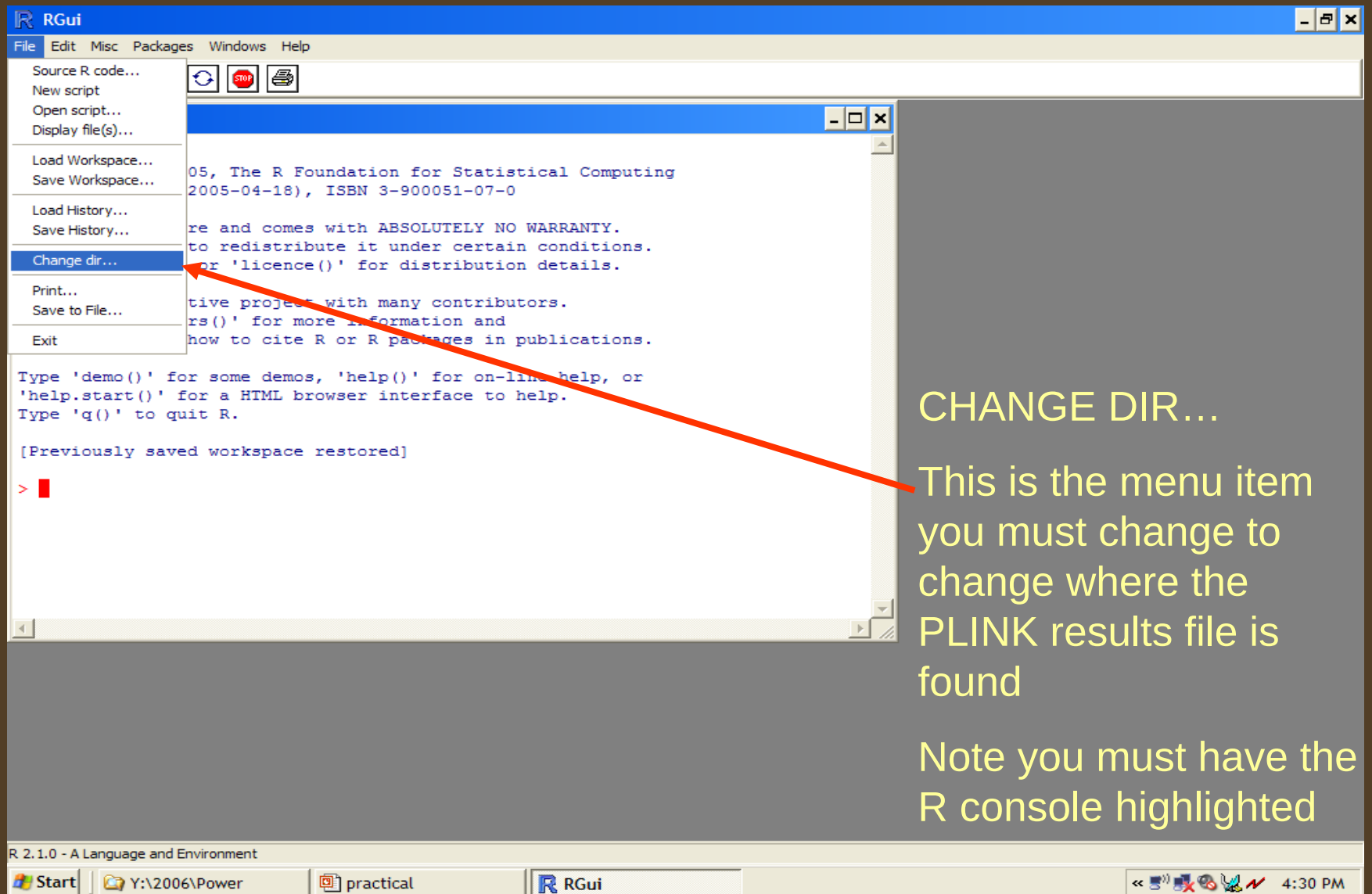
QQ plotting

Observation



Expectation

Plotting the results in R

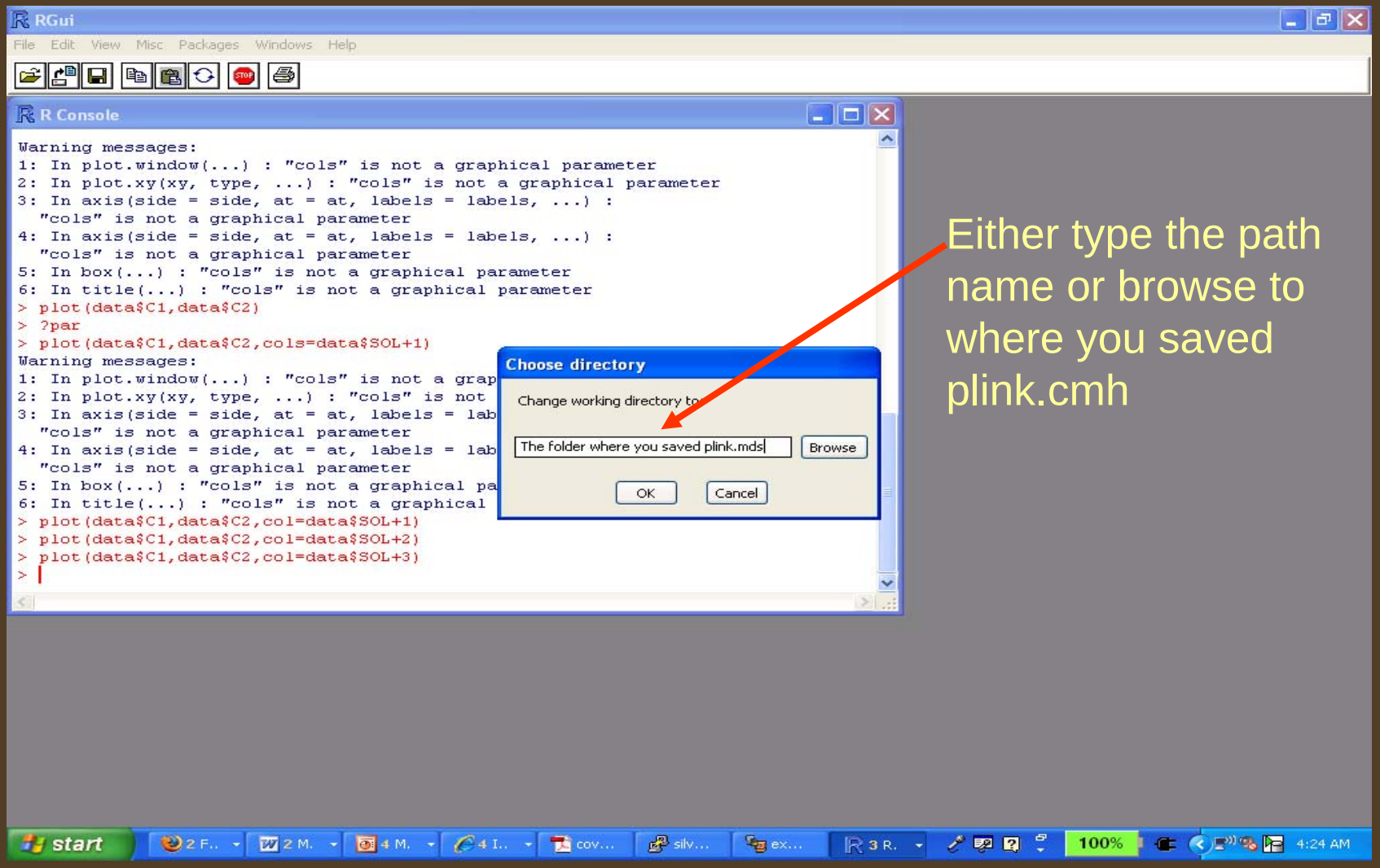


CHANGE DIR…

This is the menu item you must change to change where the PLINK results file is found

Note you must have the R console highlighted

Picture of the dialog box



Either type the path name or browse to where you saved plink.cmh

Screenshot of source code selection